

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 11, 2003, 17:31:56 ; Search time 2436 Seconds  
(without alignments)  
11003.154 Million cell updates/sec

Title: US-09-308-397-1  
Perfect score: 921  
Sequence: 1 atgactaaacagccttttt.....tagcacttttagaaaaatag 921

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*\*

- 1: gb.ba:\*\*
- 2: gb.htg:\*\*
- 3: gb.in:\*\*
- 4: gb.cm:\*\*
- 5: gb.ov:\*\*
- 6: gb.pat:\*\*
- 7: gb.ph:\*\*
- 8: gb.pl:\*\*
- 9: gb.pr:\*\*
- 10: gb.ro:\*\*
- 11: gb.sts:\*\*
- 12: gb.sy:\*\*
- 13: gb.un:\*\*
- 14: gb.vi:\*\*
- 15: em.ba:\*\*
- 16: em.fun:\*\*
- 17: em.hum:\*\*
- 18: em.in:\*\*
- 19: em.mu:\*\*
- 20: em.cm:\*\*
- 21: em.or:\*\*
- 22: em.ov:\*\*
- 23: em.pat:\*\*
- 24: em.ph:\*\*
- 25: em.pl:\*\*
- 26: em.ro:\*\*
- 27: em.sts:\*\*
- 28: em.un:\*\*
- 29: em.vi:\*\*
- 30: em.htg.hum:\*\*
- 31: em.htg.inrv:\*\*
- 32: em.htg.other:\*\*
- 33: em.htg.mu:\*\*
- 34: em.htg.pln:\*\*
- 35: em.htg.rod:\*\*
- 36: em.htg.mam:\*\*
- 37: em.htg.vrt:\*\*
- 38: em.sy:\*\*
- 39: em.htgo.hum:\*\*
- 40: em.htgo.mu:\*\*
- 41: em.htgo.other:\*\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	921	100.0	111135	2	SPNEU1906	AL449928 Streptoco
2	913	99.1	921	6	AX194049	AX194049 Sequence
3	913	99.1	11523	1	AF197933	AF197933 Streptoco
4	913	99.1	12039	1	AE007354	AE007354 Streptoco
c	913	99.1	19702	6	BD003687	BD003687 Polynucle
5	913	99.0	10925	1	AE008418	AE008418 Streptoco
6	911.4	99.0	10925	1	AE008418	AE008418 Streptoco
7	908.4	98.6	945	6	AX194275	AX194275 Sequence
c	523.2	56.8	1196	6	AR193756	AR193756 Sequence
8	439.4	47.7	10909	1	AE006603	AE006603 Streptoco
9	439.4	47.7	50416	1	AE014165	AE014165 Streptoco
c	437.8	47.5	14456	1	AE010088	AE010088 Streptoco
10	388.8	42.2	10701	1	AE006311	AE006311 Lactococc
11	388.8	42.2	10701	1	AE006311	AE006311 Lactococc
12	219.6	23.8	14530	1	AE004276	AE004276 Vibrio ch
c	212.4	23.1	11845	1	AE013105	AE013105 Thermoana
13	204	22.1	950	6	AX432728	AX432728 Sequence
14	204	22.1	950	6	AX432728	AX432728 Sequence
15	204	22.1	950	6	AX432728	AX432728 Sequence
16	204	22.1	950	6	AX432728	AX432728 Sequence
17	200.2	21.7	298050	1	AP003189	AP003189 Clostridi
c	196.4	21.3	10857	1	U32701	U32701 Haemophilus
18	187.2	20.3	12890	1	AE005319	AE005319 Escherich
19	187.2	20.3	12890	1	AE005319	AE005319 Escherich
20	187.2	20.3	222605	1	AF002555	AF002555 Escherich
21	185.6	20.2	1440	1	ECFABHDG	Z11565 E.coli fabH
22	185.6	20.2	1608	1	ECOFABD	M87040 Escherichia
23	185.6	20.2	13051	1	AE000210	AE000210 Escherich
24	185.6	20.2	16379	1	D90745	D90745 Escherichia
25	179.6	19.5	27779	1	BSV13937	Y13937 Bacillus su
26	179.6	19.5	208780	1	BSUB0009	Z93112 Bacillus su
27	179.4	19.5	10684	1	AE006228	AE006228 Pasteurel
c	178.4	19.4	13165	1	AE007854	AE007854 Clostridi
28	170	18.5	6394	1	AE0021978	AB021978 Vibrio ma
29	168	18.2	4430	1	AF044668	AF044668 Salmonell
30	168	18.2	4430	1	AF044668	AF044668 Salmonell
31	168	18.2	20097	1	AE008752	AE008752 Salmonell
c	165.8	18.0	230249	1	AP001515	AP001515 Bacillus
32	165.8	18.0	230249	1	AP001515	AP001515 Bacillus
33	165.8	18.0	230249	1	AP001515	AP001515 Bacillus
34	165	17.9	13099	1	AE013778	AE013778 Yersinia
35	165	17.9	193050	1	AJ414149	AJ414149 Yersinia
36	164.8	17.9	254050	1	AL627269	AL627269 Salmonell
c	163.4	17.7	110000	2	AC020884	Continuation (2 of
37	163.4	17.7	110000	2	AC020884	Continuation (2 of
38	162.4	17.6	313450	1	AL596170	AL596170 Listeria
c	162.4	17.6	313450	1	AL596170	AL596170 Listeria
39	162.4	17.6	313450	1	AL596170	AL596170 Listeria
c	162.4	17.6	313450	1	AL596170	AL596170 Listeria
40	162.4	17.6	34980	6	AX417046	AX417046 Sequence
c	160.4	17.4	347050	1	AL591981	AL591981 Listeria
41	160.4	17.4	347050	1	AL591981	AL591981 Listeria
42	160.2	17.4	2678	1	BSU59433	U59433 Bacillus su
43	147.4	16.0	6941	1	AE010528	AE010528 Fusobacte
44	139.6	15.2	592	6	AX437137	AX437137 Sequence
45	136.6	14.8	14534	1	AE002541	AE002541 Neisseria

ALIGNMENTS

RESULT 1	SPNEU1906	111135 bp	DNA	linear	HTG 11-JUL-2001
LOCUS	Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***				
DEFINITION	in ordered pieces.				
ACCESSION	AL449928				
VERSION	AL449928.1	GI:11545153			
KEYWORDS	HTG; HTGS PHASE2.				
SOURCE	Streptococcus pneumoniae.				
ORGANISM	Streptococcus pneumoniae				
REFERENCE	Bacteria; Firmicutes; Lactobacilliales; Streptococcaceae;				
AUTHORS	Streptococcus.				
	1 (bases 1 to 111135)				
	Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Humbert,Y.,				
	Friedl,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de				

Francisco, M., Polissi, A., Buell, G., Feger, G., Garcia, E., Feitsch, M., and Garcia-Bustos, J.F.  
Annotated draft genomic sequence from a *Streptococcus pneumoniae* type 19F clinical isolate

Microb. Drug Resist. 7 (2), 99-125 (2001)

2135329

11442348

2 (bases 1 to 111135)

Dopazo, J., Mendoza, A., Herrero, J., Caldara, F., Polissi, A., Humbert, Y., Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de Francisco, M., Buell, G., Feger, G., Garcia, E., Feitsch, M. and Garcia-Bustos, J.F.

Direct Submission

Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A., Severo Ochoa 2, 28760 Tres Cantos, SPAIN

\* NOTE: This is a 'working draft' sequence.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

FEATURES

source

1. .111135

/organism="Streptococcus pneumoniae"

/serotype="19F"

/db\_xref="taxon:1313"

/clone="G54"

BASE COUNT 32875 a 20454 c 25072 g 32680 t 54 others

ORIGIN

Query Match 100.0%; Score 921; DB 2; Length 111135;

Best Local Similarity 100.0%; Pred. No. 1.2e-241;

Matches 921; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACTAAACAGCCCTTTTATCTGCTCAAGTGCCAGTACTAGGATGGGCGG 60

Db 26478 ATGACTAAACAGCCCTTTTATCTGCTCAAGTGCCAGTACTAGGATGGGCGG 26537

QY 61 GATTTCATGATCAGTATCCGATGCTCAAGAAACGATTGATCGAGCGAGTGCCTC 120

Db 26538 GATTTCATGATCAGTATCCGATGCTCAAGAAACGATTGATCGAGCGAGTGCCTC 26597

QY 121 GGTATGATTACGTTACTCTCATGATAGGAGAGCAACTCAATCAGCCGCTAT 180

Db 26598 GGTATGATTACGTTACTCTCATGATAGGAGAGCAACTCAATCAGCCGCTAT 26657

QY 181 ACGCAACCAAGCCATCTAGCAGTCTGGTGTCTATCCGTTTATTCGAAGAAAGGCG 240

Db 26658 ACGCAACCAAGCCATCTAGCAGTCTGGTGTCTATCCGTTTATTCGAAGAAAGGCG 26717

QY 241 TATCAGCCTGATATGGTGTCTGTCTCTTGAGAAATCTCTGCTTGGTGGCAAGC 300

Db 26718 TATCAGCCTGATATGGTGTCTGTCTCTTGAGAAATCTCTGCTTGGTGGCAAGC 26777

QY 301 GGCGCCTTGGATTTGAAGATCGGTTGCTTGGTAGCTAAGCGTGGAGCCTATATGAA 360

Db 26778 GGCGCCTTGGATTTGAAGATCGGTTGCTTGGTAGCTAAGCGTGGAGCCTATATGAA 26837

QY 361 GAAGCGCTCTGCTGATCTCGCAAGATGGTAGCAGTTCTCAATACGCCAGTAGAGTC 420

Db 26838 GAAGCGCTCTGCTGATCTCGCAAGATGGTAGCAGTTCTCAATACGCCAGTAGAGTC 26897

QY 421 ATTGAAGAGCGCTGTCAAAGACCTTCTGAACTTGGAGTGGTTACTCCAGGCCAATATAAC 480

Db 26898 ATTGAAGAGCGCTGTCAAAGACCTTCTGAACTTGGAGTGGTTACTCCAGGCCAATATAAC 26957

QY 481 ACACCTGCAAAATCGTANTCTCGAGAGATGGTTGAGTTGATCGAGCGGTTGAATC 540

Db 26958 ACACCTGCAAAATCGTANTCTCGAGAGATGGTTGAGTTGATCGAGCGGTTGAATC 27017

QY 541 TTGCAAGAGAGCGGTGCCAAAGCTTGTCTCTTAAGGTGTCAGGTCCCTTTTCACAC 600

Db 27018 TTGCAAGAGAGCGGTGCCAAAGCTTGTCTCTTAAGGTGTCAGGTCCCTTTTCACAC 27077

QY 601 GCTCTCTTGGCCCTGCTAGCCAGAACTAGCTGAAACTCTAGCTCAGGTAAGTTTTC 660

Db 27078 GCTCTCTTGGCCCTGCTAGCCAGAACTAGCTGAAACTCTAGCTCAGGTAAGTTTTC 27137

QY 661 GATTTCATGATCAGTATCCGATGCTCAAGAAACGATTGATCGAGCGAGTGCCTC 720

Db 27138 GATTTCATGATCAGTATCCGATGCTCAAGAAACGATTGATCGAGCGAGTGCCTC 27197

QY 721 GCTCAGCTCTTGACCGCTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTC 780

Db 27198 GCTCAGCTCTTGACCGCTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTC 27257

QY 781 ATGCAAGAGCAGCGCATTAAGCAACTTTATCGAGATTGACCGGGGAAAGTCTTTGTCAGGT 840

Db 27258 ATGCAAGAGCAGCGCATTAAGCAACTTTATCGAGATTGACCGGGGAAAGTCTTTGTCAGGT 27317

QY 841 TTGTTTAAAAAATTCATCAAACTGCTCACTTAGCTCATGTGGAAGATCAAGCGAGTTTA 900

Db 27318 TTGTTTAAAAAATTCATCAAACTGCTCACTTAGCTCATGTGGAAGATCAAGCGAGTTTA 27377

QY 901 GTAGCAGCTTTTAAAAAATAG 921

Db 27378 GTAGCAGCTTTTAAAAAATAG 27398

RESULT 2

AX194049

LOCUS

AX194049

DEFINITION

Sequence 30 from Patent WO0149721..

ACCESSION

AX194049

VERSION

AX194049.1

GI:15211644

KEYWORDS

Streptococcus pneumoniae.

ORGANISM

Streptococcus pneumoniae

Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

Streptococcus.

REFERENCE

1 (bases 1 to 921)

Dougherty, T.J., Pucci, M.J., Dougherty, B.A., Davison, D.B.,

Brucoleri, R.E. and Thanassi, J.A.

Novel bacterial genes and proteins that are essential for cell

viability and their uses

JOURNAL

Patent: WO 0149721-A 30 12-JUL-2001;

Bristol-Myers Squibb Co. (US)

FEATURES

Location/Qualifiers

1..921

/organism="Streptococcus pneumoniae"

/db\_xref="taxon:1313"

BASE COUNT 246 a 193 c 232 g 250 t

ORIGIN

Query Match 99.1%; Score 913; DB 6; Length 921;

Best Local Similarity 99.5%; Pred. No. 1.6e-239;

Matches 916; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ATGACTAAACAGCCCTTTTATCTGCTCAAGTGCCAGTACTAGGATGGGCGG 60

Db 1 ATGACTAAACAGCCCTTTTATCTGCTCAAGTGCCAGTACTAGGATGGGCGG 60

QY 61 GATTTCATGATCAGTATCCGATGCTCAAGAAACGATTGATCGAGCGAGTGCCTC 120

Db 61 GATTTCATGATCAGTATCCGATGCTCAAGAAACGATTGATCGAGCGAGTGCCTC 120

QY 121 GGTATGATTACGTTACTCTCATGATAGGAGAGCAACTCAATCAGACCCGCTAT 180

Db 121 GGTATGATTACGTTACTCTCATGATAGGAGAGCAACTCAATCAGACCCGCTAT 180

QY 181 ACGCAACCAAGCCATCTAGCAGTCTGGTGTCTATCTACCGTTTATTCGAAGAAAGGCG 240

Db 181 ACGCAACCAAGCCATCTAGCAGTCTGGTGTCTATCTACCGTTTATTCGAAGAAAGGCG 240

QY 241 TATCAGCCTGATATGGTGTCTGTCTCTTGAGAAATCTCTGCTTGGTGGCAAGC 300

Db 241 TATCAGCCTGATATGGTGTCTGTCTCTTGAGAAATCTCTGCTTGGTGGCAAGC 300

```
QY 301 GGCGCTTGGATTTTGAAGATGCGGTGCGCTTGTGCTAGCTAAGCGGTGAGCCTATATGGAA 360
Db 301 GGCGCTTGGATTTTGAAGATGCGGTGCGCTTGTGCTAGCTAAGCGGTGAGCCTATATGGAA 360
QY 361 GAAGCGGCTCCTGCTGACTCTGCGCAAGATGTTAGCAGTTCTCAATACGCCAGTAGAGGTC 420
Db 361 GAAGCGGCTCCTGCTGACTCTGCGCAAGATGTTAGCAGTTCTCAATACGCCAGTAGAGGTC 420
QY 421 ATTGAAGAGCCTGTCAAAAGCTTCTGAACCTTGAGCTTGAGTGGTGTACTCCAGCCTATATAC 480
Db 421 ATTGAAGAGCCTGTCAAAAGCTTCTGAACCTTGAGTGGTGTACTCCAGCCTATATAC 480
QY 481 ACACCTGCACAAATCGCTATTGCTGGAGAGTGGTTCGAGTTGATCGAGCGGTTGAACCTT 540
Db 481 ACACCTGCACAAATCGCTATTGCTGGAGAGTGGTTCGAGTTGATCGAGCGGTTGAACCTT 540
QY 541 TTGCAAGAGCAGGTGCGCAAGCTTGAATTCCTTTAAGTGTGAGTGCAGTCCCTTTCAACC 600
Db 541 TTGCAAGAGCAGGTGCGCAAGCTTGAATTCCTTTAAGTGTGAGTGCAGTCCCTTTCAACC 600
QY 601 GCTCTCCTTGAGCTGTAGCGCAAACTAGTCAGAACTCTAGCTCAGGTAAGTTTCA 660
Db 601 GCTCTCCTTGAGCTGTAGCGCAAACTAGTCAGAACTCTAGCTCAGGTAAGTTTCA 660
QY 661 GATTTTACTTGTCCCTTAGTGGCAATACAGAACTCTGTGTGATGCAAAAAGAGGACATT 720
Db 661 GATTTTACTTGTCCCTTAGTGGCAATACAGAACTCTGTGTGATGCAAAAAGAGGACATT 720
QY 721 GCTCAGCTCTTGAGCGCTCAGGTCAAGGACCGGTCGTTCTTATGAAAGTATTTGGGTC 780
Db 721 GCTCAGCTCTTGAGCGCTCAGGTCAAGGACCGGTCGTTCTTATGAAAGTATTTGGGTC 780
QY 781 ATGCAAGAGCAGGCATAGCAACTTTATCGAGTTGCGACGGGAAAGCTTTGTCAGGT 840
Db 781 ATGCAAGAGCAGGCATAGCAACTTTATCGAGTTGCGACGGGAAAGCTTTGTCAGGT 840
QY 841 TTTGTTAAAAAATTGATCAACTGCTCAGTCTAGCTCATGTGGAAGTCAAGCAGTTTA 900
Db 841 TTTGTTAAAAAATTGATCAACTGCTCAGTCTAGCTCATGTGGAAGTCAAGCAGTTTA 900
QY 901 GTAGCAGCTTTAGAAAAATAG 921
Db 901 GTAGCAGCTTTAGAAAAATAG 921

RESULT 3
AF197933 11523 bp DNA linear BCT 10-AUG-2000
LOCUS Streptococcus pneumoniae fab gene cluster, complete sequence.
DEFINITION AF197933
ACCESSION AF197933
VERSION AF197933.1 GI:9789228
KEYWORDS
SOURCE Streptococcus pneumoniae.
ORGANISM Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
1 (bases 1 to 11523)
Heath,R.J. and Rock,C.O.
A triclozan-resistant bacterial enzyme
JOURNAL Nature 406 (6792), 145-146 (2000)
MEDLINE 20365714
PUBMED 10910344
REFERENCE 2 (bases 1 to 11523)
Heath,R.J. and Rock,C.O.
Direct Submission
JOURNAL Submitted (25-OCT-1999) Department of Biochemistry, St Jude
Children's Research Hospital, 332 North Lauderdale Street, Memphis,
TN 38105, USA
FEATURES
source
1. .11523
/organism="Streptococcus pneumoniae"
/strain="R6"
/db_xref="taxon:1313"
```

```
gene
CDS
```

```
210. .1184
/gene="fabH"
210. .1184
/gene="fabH"
/note="fabH; KASIII; acetoacetyl-ACP synthase; by
similarity to FABD_ECOLI (39% identity)"
/codon_start=1
/transl_table=11
/product="beta-ketoacyl-ACP synthase III"
/protein_id="AAF98271.1"
/db_xref="GI:9789229"
/translation="MAFAKISQVAHYVEQVVTNHLAQIMDTNDEMSSRTGIRORH
LSRSTSDLAETAVAKLMKAGITGELDFILATITPDMSMPSTAAARVOANIGANK
AFAPLTAACSGFVFALTAFTASCFQKGLVGTSETLSKAVDMSKSTAVLPGDG
AGCVLLASGEHFLAESLNSDGRSECLTYGSHGLSPFSDQSDASDFLKMGRVTF
DFAPLVAKSINKITIDESPIEVTDLDDYLLHQANDRIIDKMARKIGVDRAPANNME
YGNSTAASIPILLSECVEQGLIPLDGSCTVLLSGFGLTWKTLILTI"
1244. .1468
/gene="acpP"
1244. .1468
/gene="acpP"
/note="ACP; by similarity to ACP_P_ECOLI (38% identity)"
/codon_start=1
/transl_table=11
/product="acyl carrier protein"
/protein_id="AAF98272.1"
/db_xref="GI:9789230"
/translation="MAVPEKVOELIVEELGKDASEVTLESTFDDLDADSLDFQVISE
TEDAPDQIEAENDLKTGVDLVAIVEEQAK"
1587. .2561
/gene="fabK"
1587. .2561
/gene="fabK"
/Note="FabK; NADH-dependent, FAD-containing enoyl-ACP
reductase"
/codon_start=1
/evidence="experimental"
/transl_table=11
/product="trans-2-enoyl-ACP reductase II"
/protein_id="AAF98273.1"
/db_xref="GI:9789231"
/translation="MKTFITELLKIDVPIFOGMAVADGLAGAVSKAGGLIGGG
NAPKEVXANIDKIKSLTDKPGVYNIMLLSPFVEILDVIEEGVKVVTTCGAPNSKY
MERFHEAGITVIVPVSVALKRMEKADAVIAGNEAGGHIKLTMTLVRQVATA
ISIPIVIAAGGIADGGAAGMGAEVQVGTFRVAKESNAHNPKEKILKARDIDT
TISAGHGHAVRAIKNQLTRDFELAKEDAFQEDPDLEIFQMGAGALAKAVHGDVD
GGSVMAGQIAGLSKETAETAEILKDLXYGAARKIQEASRWTVVRND"
2554. .3474
/gene="fabD"
2554. .3474
/gene="fabD"
/Note="FabD; by similarity to FABD_ECOLI (42% identity);
MCAT"
/codon_start=1
/transl_table=11
/product="malonyl-CoA:ACP transacylase"
/protein_id="AAF98274.1"
/db_xref="GI:9789232"
/translation="MTKTFALFAGGQAQYLGMRDFDYDQYPIVKTETDRASQVLGYDL
RYLIDTEEDKLNQRYTQPAITATSVAIYRLLOKGYOPDMVAGLSGEYSALVASGA
LDFEDAVALKRGAYMEAPADSGKMWAVINTFVEIEEACQKSELGVYTPANNY
TPAQIVAGEVVAVDRAVLEQAGAKRLIPLKVSPPHTALLEPASOKLAETLAQVS
PSDFTCPVGNTEAAVMQKEDIAQLLTQVKEPVRFVYESIGVMQEAGISNFIETPGPK
VLSGFVKKIDQTAHLAHVEDQASLVALLEK"
3508. .4239
/gene="fabG"
3508. .4239
/gene="fabG"
/Note="FabG; by similarity to FABG_ECOLI (44% identity)"
/codon_start=1
/transl_table=11
/product="beta-ketoacyl-ACP reductase"
/protein_id="AAF98275.1"
```

```
gene
CDS
```

```
210. .1184
/gene="fabH"
210. .1184
/gene="fabH"
/note="fabH; KASIII; acetoacetyl-ACP synthase; by
similarity to FABD_ECOLI (39% identity)"
/codon_start=1
/transl_table=11
/product="beta-ketoacyl-ACP synthase III"
/protein_id="AAF98271.1"
/db_xref="GI:9789229"
/translation="MAFAKISQVAHYVEQVVTNHLAQIMDTNDEMSSRTGIRORH
LSRSTSDLAETAVAKLMKAGITGELDFILATITPDMSMPSTAAARVOANIGANK
AFAPLTAACSGFVFALTAFTASCFQKGLVGTSETLSKAVDMSKSTAVLPGDG
AGCVLLASGEHFLAESLNSDGRSECLTYGSHGLSPFSDQSDASDFLKMGRVTF
DFAPLVAKSINKITIDESPIEVTDLDDYLLHQANDRIIDKMARKIGVDRAPANNME
YGNSTAASIPILLSECVEQGLIPLDGSCTVLLSGFGLTWKTLILTI"
1244. .1468
/gene="acpP"
1244. .1468
/gene="acpP"
/Note="ACP; by similarity to ACP_P_ECOLI (38% identity)"
/codon_start=1
/transl_table=11
/product="acyl carrier protein"
/protein_id="AAF98272.1"
/db_xref="GI:9789230"
/translation="MAVPEKVOELIVEELGKDASEVTLESTFDDLDADSLDFQVISE
TEDAPDQIEAENDLKTGVDLVAIVEEQAK"
1587. .2561
/gene="fabK"
1587. .2561
/gene="fabK"
/Note="FabK; NADH-dependent, FAD-containing enoyl-ACP
reductase"
/codon_start=1
/evidence="experimental"
/transl_table=11
/product="trans-2-enoyl-ACP reductase II"
/protein_id="AAF98273.1"
/db_xref="GI:9789231"
/translation="MKTFITELLKIDVPIFOGMAVADGLAGAVSKAGGLIGGG
NAPKEVXANIDKIKSLTDKPGVYNIMLLSPFVEILDVIEEGVKVVTTCGAPNSKY
MERFHEAGITVIVPVSVALKRMEKADAVIAGNEAGGHIKLTMTLVRQVATA
ISIPIVIAAGGIADGGAAGMGAEVQVGTFRVAKESNAHNPKEKILKARDIDT
TISAGHGHAVRAIKNQLTRDFELAKEDAFQEDPDLEIFQMGAGALAKAVHGDVD
GGSVMAGQIAGLSKETAETAEILKDLXYGAARKIQEASRWTVVRND"
2554. .3474
/gene="fabD"
2554. .3474
/gene="fabD"
/Note="FabD; by similarity to FABD_ECOLI (42% identity);
MCAT"
/codon_start=1
/transl_table=11
/product="malonyl-CoA:ACP transacylase"
/protein_id="AAF98274.1"
/db_xref="GI:9789232"
/translation="MTKTFALFAGGQAQYLGMRDFDYDQYPIVKTETDRASQVLGYDL
RYLIDTEEDKLNQRYTQPAITATSVAIYRLLOKGYOPDMVAGLSGEYSALVASGA
LDFEDAVALKRGAYMEAPADSGKMWAVINTFVEIEEACQKSELGVYTPANNY
TPAQIVAGEVVAVDRAVLEQAGAKRLIPLKVSPPHTALLEPASOKLAETLAQVS
PSDFTCPVGNTEAAVMQKEDIAQLLTQVKEPVRFVYESIGVMQEAGISNFIETPGPK
VLSGFVKKIDQTAHLAHVEDQASLVALLEK"
3508. .4239
/gene="fabG"
3508. .4239
/gene="fabG"
/Note="FabG; by similarity to FABG_ECOLI (44% identity)"
/codon_start=1
/transl_table=11
/product="beta-ketoacyl-ACP reductase"
/protein_id="AAF98275.1"
```

```
/db xref="GI:9789233"
/translation="MKLEHKNFITGSSRGIGLAIHAKPAQAAGNIVINSRGAISEEL
LAEFNYGKIVPISGVSDPDAKRMIDQAEIGSVDDVLVNNAGITQDITMLKMTTE
ADPEKVLNLTGAFNMTQSVLKPMMKAREGAIINNSVYGLMGNIGNAYAAKAGL
IGFTKSVAREASIRNVIAPGMIESDMTAILSDKIEATLAQIPMKERFQAQBOVA
DLTVFLAGQDYLITGVVAIDGGLSM"
gene
CDS
4261..5496
/genes="fabF"
/genes="fabF"
4261..5496
/genes="fabF"
/notes="FabF; KASII; oxoacyl-ACP synthase II; by similarity
to FABF_ECOLI (43% identity) and FABF_ECOLI (33%
identity)"
/codon_start=1
/transl_table=11
/product="beta-ketoacyl-ACP synthase II"
/protein_id="AAF98276.1"
/db xref="GI:9789234"
/translation="MKLNRVVVTGYGVTSPIGNTEEFWNSLATKIGIGITKPDHS
DFVHNAEIQPFDPDKYFKKEDTNRFDNYSLYALAAQEA VNHANLDEALNDRFG
VIVASGIGIKIEIDQVLEHKEGPKRVKPMTLPRALPNMAGNVAMRFGANGVCKSI
NTACSSNDAIGDAPRSIKFGQDVLVGGTEASITPEALAGFQALFALSTEDPTRA
SIPFDKDRNGFVWGGSGMLVLESLEHAEKRGATILAEVVGVCNTCDAYHMTSPHPEG
QQAIIKALKALAEASISPOVAYVNAHGISTFANEKRGSAIVAVLGRVFPVSTKSF
TGHLLGAGAAVEAIVTIEAMRNHFVPMTAGTSEVSDYIEANVYQGLEKEIIPAYISN
TFGFGHNAVLAFKWEER"
gene
CDS
5499..5984
/genes="accB"
/genes="accB"
5499..5984
/genes="accB"
/notes="BCCP; by similarity to BCCP_ECOLI (37% identity);
subunit of acetyl-CoA carboxylase"
/codon_start=1
/transl_table=11
/product="biotin carboxyl carrier protein"
/protein_id="AAF98277.1"
/db xref="GI:9789235"
/translation="KNLNDIKDLMTQFDOSSLRERSYKNGTDELOFSKNEARPYEVA
TOVAPAFVLTSPVAPTSAPAEVPAEASVATEGLNLSPLVGVVYLAAGP
DKPAFVTVGDSVKKGOTLIIEMKVMNBIAPKDGVTTEILVSNSEEMVERFGKLVRI
K"
gene
CDS
5981..6403
/genes="fabZ"
/genes="fabZ"
5981..6403
/genes="fabZ"
/notes="FabZ; by similarity to FABZ_ECOLI (46% identity)"
/codon_start=1
/transl_table=11
/product="beta-hydroxyacyl-ACP dehydratase"
/protein_id="AAF98278.1"
/db xref="GI:9789236"
/translation="MIDIQIKELPHRYPMILLVDRVLEVS EDTIVAIKNVTINEPFF
NGHFPQYPMPGVVMEALQATAGVLELSKPKENKGLVIFYAGMDKVRKQVPODQL
VMTATFVKRGITAVVEAKAEVDGKLAASGTLTFAIGN"
gene
CDS
6415..7782
/genes="accC"
/genes="accC"
6415..7782
/genes="accC"
/notes="AccC; by similarity to ACCC_ECOLI (53% identity)"
/codon_start=1
/transl_table=11
/product="acetyl-CoA carboxylase biotin carboxylase
subunit"
/protein_id="AAF98279.1"
/db xref="GI:9789237"
/translation="MFKILIANRGEIIVRIIRAREBELGIATVAVYSTADKEALHTLL
ADEAVICGPKGATSLNINAVLSAULTEAEAIHPFGFUSENSKFMTCVEEYKIF
IGPSGHVMDMGDKINARAQMIKAGVPVPGSDGEVHNSSEALIVAEKIGYPMVKAS
AGGGKGIKVKPDLVSAFETASSEAKANYNGAMYIERVYPARHIEYQILGDEH
GWHILGERDCSLQRNNQKLVLESPSIAIKTLRHEIGAAAVRAAEFVYENAGTIEF
LIDEASNFYFMEMTRVOVEHPVTEFVSGVDIVKEQICIAAGOPLSVKQDILVRGH
AIECRINAPNAPFAPSFGKITNLVLPSSGVLGVDSAVYGVGTIPYVDSMLAKII
VHGENRFDLMCKQALIELEGVQTNADFQDLDISRNVIAGDYDYSFLMEYFLPK
```

```
YQEKKE"
7819..8885
/genes="accD"
7819..8885
/genes="accD"
/notes="AccD; by similarity to ACCD_ECOLI (43% identity)"
/codon_start=1
/transl_table=11
/product="acetyl-CoA carboxylase beta subunit"
/protein_id="AAF98280.1"
/db xref="GI:9789238"
/translation="MALFSKKDKYIRINPNRSVREKPOAKPEVDFELFQCPGCKHTI
YKDLGSRICPHCSYTFPISAOERLALIDMGTFKELFTG:ESKQPLHFPYQKKLA
SMREKTLGHEAVVTGALIKGOTVALIGMDSFMASMGTVVGEKILTRLFVATVKLA
PVVLFTASCGARMQEGIMSLQMAKISAAVKSHSNAGLYLAILTPTTGGVYASAM
EGDILIAEQSLVGFAGRRVIENTVRESLPEDFKAEFLLEHGFVDIVARRDLPTTI
ASLVRHGGSPR"
gene
CDS
8682..9449
/genes="accA"
/genes="accA"
8682..9449
/genes="accA"
/notes="AccA; by similarity to ACCA_ECOLI (46% identity)"
/codon_start=1
/transl_table=11
/product="acetyl-CoA carboxylase alpha subunit"
/protein_id="AAF98281.1"
/db xref="GI:9789239"
/translation="MNIKIVREAREQSRLTLTDFATGIFDPEFIQLHGDSPRDDGAV"
Query Match 99.1%; Score 913; DB 1; Length 11523;
Best Local Similarity 99.5%; Pred. No. 1.8e-239;
Matches 916; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 1 ATGACTAAACACAGCCCTTTTATTTGCTGTGCTCAAGGTGCCAGTATCTAGGGATGGACGG 60
Db 2554 ATGACTAAACACAGCCCTTTTATTTGCTGTGCTCAAGGTGCCAGTATCTAGGGATGGACGG 2613
Qy 61 GATTTCTATGATCAGTATCCGATTGTCAAGAAACGATTGATCGAGGAGTCAGGTGCTC 120
Db 2614 GATTTCTATGATCAGTATCCGATTGTCAAGAAACGATTGATCGAGGAGTCAGGTGCTC 2673
Qy 121 GGTATGATTTAGTTCATCTCATGATACGGAAGAAGACAACTCAATCAGACCCGCTAT 180
Db 2674 GGTATGATTTAGTTCATCTCATGATACGGAAGAAGACAACTCAATCAGACCCGCTAT 2733
Qy 181 ACGCAACACGCCATTCCTAGCGATTCGGTTGCTATCTACCGTTTATTCGAAAGAAAGGCG 240
Db 2734 ACGCAACACGCCATTCCTAGCGATTCGGTTGCTATCTACCGTTTATTCGAAAGAAAGGCG 2793
Qy 241 TATCAGCCTCATATGGTTGCTGTTGCTCTCTTGGAGATACCTCTGCTTGGTGGCAAGC 300
Db 2794 TATCAGCCTCATATGGTTGCTGTTGCTCTCTTGGAGATACCTCTGCTTGGTGGCAAGT 2853
Qy 301 GCGCCCTTGGATTTGAAGATCGGTTGCTGCTAGTAAAGCGTGGAGCCTATATGGAA 360
Db 2854 GCGCCCTTGGATTTGAAGATCGGTTGCTGCTAGTAAAGCGTGGAGCCTATATGGAA 2913
Qy 361 GAAGCGCTCCTGCTGACTCTGCAAGATGGTAGAGTTCTCAATACGCCAGTAGAGGTC 420
Db 2914 GAAGCGCTCCTGCTGACTCTGCAAGATGGTAGAGTTCTCAATACGCCAGTAGAGGTC 2973
Qy 421 ATTGAAGAAGCCCTGTCAAAAGAGCTTCTGAACCTTGGAGTGGTTACTCCAGCCAACTATAAC 480
Db 2974 ATTGAAGAAGCCCTGTCAAAAGAGCTTCTGAACCTTGGAGTGGTTACTCCAGCCAACTATAAC 3033
Qy 481 ACACCTGCACAAATCGTCATCTGCGAAGATGGTTGCGAGTTGATCGAGCGGTGTAACCT 540
Db 3034 ACACCTGCACAAATCGTCATCTGCGAAGATGGTTGCGAGTTGATCGAGCGGTGTAACCT 3093
Qy 541 TTGCAAGAAGCAGGTGCCAAACGCTTGATTCCTCTTAAAGTGTGTCAGGTCCCTTTTCACAC 600
Db 3094 TTGCAAGAAGCAGGTGCCAAACGCTTGATTCCTCTTAAAGTGTGTCAGGTCCCTTTTCACAC 3153
```





```
QY 901 GTAGCAGCTTTAGAAAAATAG 921
Db 1957 GTAGCAGCTTTAGAAAAATAG 1977

RESULT 5
LOCUS BD003687/c 19702 bp DNA linear PAT 31-JAN-2002
DEFINITION Polynucleotide of Streptococcus pneumoniae and sequence.
ACCESSION BD003687
VERSION BD003687.1 GI:18631648
KEYWORDS JP 2001501833-A/7.
SOURCE unclassified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 19702)
AUTHORS Kunsch,C.A., Choi,G.H., Dillon,P.J., Rosen,C.A., Bara,S.C.,
Fannon,M. and Dougherty,B.A.
TITLE Polynucleotide of Streptococcus pneumoniae and sequence
JOURNAL Patent: JP 2001501833-A 7 13-FEB-2001;
HUMAN GENOME SCIENCES INC
COMMENT OS Unidentified
PN JP 2001501833-A/7
PD 13-FEB-2001
PF 30-OCT-1997 JP 1998520718
PR 31-OCT-1996 US 60/029960
PI CHARLES A KUNSCH, GIL H CHOI, PATRICK J DILLON, CRAIG A ROSEN, PI
STEVEN C BARASH,
PI MICHAEL FANNON BRIAN A DOUGHERTY
PC C12N5/09,A01K67/027,C07K14/315,C07K16/12,C12N1/15,C12N1/19,
PC C12N1/21,
PC C12N5/10,C12P21/02,C12Q1/68,G06F17/30,C12N15/00,C12N5/00, PC
G06F15/40
CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..19702
FT location/Qualifiers
FT /organism="Unidentified".
1..19702
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 5736 a 4548 c 3728 g 5690 t
ORIGIN

Query Match 99.1%; Score 913; DB 6; Length 19702;
Best Local Similarity 99.5%; Pred. No. 1.8e-239;
Matches 916; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ATGCTTAAACAGCGCTTTTATTTGCTGCTCAAGTCCCGAGTATCTAGGGTGGGACGG 60
Db 17149 ATGCTTAAACAGCGCTTTTATTTGCTGCTCAAGTCCCGAGTATCTAGGGTGGGACGG 17090

QY 61 GATTTCATGATCATCGATATCCGATTGTCAAGAAACGATTGATCGAGCGAGTTCAGGTGCTC 120
Db 17089 GATTTCATGATCATCGATATCCGATTGTAAAGAAACGATTGATCGAGCGAGTTCAGGTGCTC 17030

QY 121 GGTTATGATTACGTTATCTCATGATACGGAAGAACAACTCAATCAGCCCGTAT 180
Db 17029 GGTTATGATTACGTTATCTCATGATACGGAAGAACAACTCAATCAGCCCGTAT 16970

QY 181 ACGCAACAGGACATTCAGCGACATTCGCTGCTATCTACCGTTTATTCGAAGAAAGGCG 240
Db 16969 ACGCAACAGGACATTCAGCGACATTCGCTGCTATCTACCGTTTATTCGAAGAAAGGCG 16910

QY 241 TATCAGCCTGATATGTTGCTGTTGCTCTTGGAGAACTCTGCTTGGTGGGCAAGC 300
Db 16909 TATCAGCCTGATATGTTGCTGTTGCTCTTGGAGAACTCTGCTTGGTGGGCAAGT 16850

QY 301 GCGCCTTGGATTGTAAGATGCGGTTGCTTGTAGTAAAGCTGAGGCTATATGGA 360
Db 16849 GCGCCTTGGATTGTAAGATGCGGTTGCTTGTAGTAAAGCTGAGGCTATATGGA 16790
```

```
QY 361 GAAGCGGCTCTGCTGACTCTGGCAAGATGGTAGCAGTTCTCAATACGCCAGTACAGGTC 420
Db 16789 GAAGCGGCTCTGCTGACTCTGGCAAGATGGTAGCAGTTCTCAATACGCCAGTACAGGTC 16730

QY 421 ATTGAAGAAGCCTGTCTCAAAAAGCTTCTGAACCTTGAGTGGTTACTCCAGCCAACTATAAC 480
Db 16729 ATTGAAGAAGCCTGTCTCAAAAAGCTTCTGAACCTTGAGTGGTTACTCCAGCCAACTATAAC 16670

QY 481 ACACCTGCAAAATCGTATCTGCGAGAAGTGGTTGCAAGTTGATCGAGCGGTTGAACCTT 540
Db 16669 ACACCTGCAAAATCGTATCTGCGAGAAGTGGTTGCAAGTTGATCGAGCGGTTGAACCTT 16610

QY 541 TTGCAAGAAGCAGGTGGCCAAACGCTTATTCTCTTAAGGTGTAGGTCCTTTTACACC 600
Db 16609 TTGCAAGAAGCAGGTGGCCAAACGCTTATTCTCTTAAGGTGTAGGTCCTTTTACACC 16550

QY 601 GCTCTCTTTGAGCGCTGCTAGCCAGAACTAGCTGAAACTTAGCTCAGTAAAGTTTTCAC 660.
Db 16549 GCTCTCTTTGAGCGCTGCTAGCCAGAACTAGCTGAAACTTAGCTCAGTAAAGTTTTCAC 16490

QY 661 GATTTTACTTGTCCCTTAGTGGCAATACAGAAGCTGTGTGATGATCAAAAAGAGGACATT 720
Db 16489 GATTTTACTTGTCCCTTAGTGGCAATACAGAAGCTGTGTGATGATCAAAAAGAGGACATT 16430

QY 721 GCTCAGCTCTTCACGCGTCAAGTCAAGAACCCGTTCTTTCTATGAAAGTATTGGGGTC 780
Db 16429 GCTCAGCTCTTCACGCGTCAAGTCAAGAACCCGTTCTTTCTATGAAAGTATTGGGGTC 16370

QY 781 ATGCAAGAGCAGGATAGCAACTTATCGAGTTGACCGGGAAGTCTTGTCAAGT 840
Db 16369 ATGCAAGAGCAGGATAGCAACTTATCGAGTTGACCGGGAAGTCTTGTCAAGT 16310

QY 841 TTGTTTAAAAAATGATCAAACTGCTCAGTCTAGCTCATGTGGAAGATCAACGAGTTTA 900
Db 16309 TTGTTTAAAAAATGATCAAACTGCTCAGTCTAGCTCATGTGGAAGATCAACGAGTTTA 16250

QY 901 GTAGCAGCTTTAGAAAAATAG 921
Db 16249 GTAGCAGCTTTAGAAAAATAG 16229

RESULT 6
RELOCUS AE008418 10925 bp DNA linear BCT 13-SEP-2001
DEFINITION Streptococcus pneumoniae R6 section 34 of 184 of the complete genome.
ACCESSION AE008418 AB007317
VERSION AE008418.1 GI:15457935
KEYWORDS Streptococcus pneumoniae R6.
SOURCE Streptococcus pneumoniae R6.
ORGANISM Streptococcus pneumoniae R6.
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.
REFERENCE 1 (bases 1 to 10925)
AUTHORS Hoskins,J.A., Alborn,W. Jr., Arnold,J., Blaszcak,L., Burgett,S., DeHoff,B.S., Estrem,S., Fritz,L., Fu,D.-J., Fuller,W., Geringer,C., Gilmore,R., Glass,J.S., Khoja,H., Kraft,A., Lagace,R., LeBlanc,D.J., Lee,L.N., Leikowitz,E.J., Lu,J., Matsushima,P., McLaren,S., McHenry,M., McLeaster,K., Mundy,C., Nicas,I.I., Norris,F.H., O'Garra,M., Peery,Y., Robertson,G.T., Rockey,P., Sun,P.-M., Winkler,M.E., Yang,Y., Young-Bellido,M., Zhao,G., Zook,C., Baltz,R.H., Jaskunas,S.R., Rostek,P.R. Jr., Skatrud,P.L. and Glaes,J.I.
TITLE Genome of the bacterium Streptococcus pneumoniae strain R6
JOURNAL J. Bacteriol. 183 (19), 5709-5717 (2001)
MEDLINE 21425245
PUBMED 11544234
REFERENCE 2 (bases 1 to 10925)
AUTHORS Hoskins,J.A., Alborn,W. Jr., Arnold,J., Blaszcak,L., Burgett,S., DeHoff,B.S., Estrem,S., Fritz,L., Fu,D.-J., Fuller,W., Geringer,C., Gilmore,R., Glass,J.S., Khoja,H., Kraft,A., Lagace,R., LeBlanc,D.J., Lee,L.N., Leikowitz,E.J., Lu,J., Matsushima,P., McLaren,S., McHenry,M., McLeaster,K., Mundy,C., Nicas,I.I., Norris,F.H., O'Garra,M., Peery,Y., Robertson,G.T., Rockey,P., Sun,P.-M., Winkler,M.E., Yang,Y., Young-Bellido,M., Zhao,G., Zook,C., Baltz,R.H., Jaskunas,S.R., Rostek,P.R. Jr., Skatrud,P.L. and Glaes,J.I.
```



KSINTACSSNDAIGDAFRSIRKGFODVNLVGGTBSAIIPTPAIAGQALTAIATSTEDP  
TRASIPDKDRNGVMEGSGMLVLESLEHAKRGATILAEVVGYNCTDAYHTSPH  
PEGGAIKAIKLAJEEAEISPEQVYVYNAHGTSTPANKEGSAIVAVLUGKEVPSST  
KSFTGHLGAGAVEAIVTIEAMRNFVPMPTAGTSEVSDYIEANVYVGGLEKEBIPYA  
ISNTFGFGHNAVLAFKRWNR"  
6866..7351  
/gene="accb"  
/note="spr0383"  
6866..7351  
/gene="accb"  
/EC\_number="6.4.1.2"  
/codon\_start=1  
/transl\_table=11  
/label="spr0383"  
/product="Biotin carboxyl carrier protein of acetyl-CoA  
carboxylase"  
/protein\_id="AAK99187.1"  
/db\_xref="GI:15457944"  
TOVAPALATPSPVAPTSAPAEVAPAEASVATEGNLVESPLGVVYLAAGP  
DKPAFTVGSVVRKQTLVIIEMKVNNEIPAPKDGVTTEILVSNEEMVEFGKGLVRI  
K"  
7348..7770  
/gene="fabz"  
Query Match 99.0%; Score 911.4; DB 1; Length 10925;  
Best Local Similarity 99.3%; Pred. No. 4.8e-239;  
Matches 915; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 1 ATGACTAAACAGCGCTTTTATTGCTGGTCAAGGTGCCAGATCTAGGGATGGACGG 60  
DB 3921 ATGACTAAACAGCGCTTTTATTGCTGGTCAAGGTGCCAGATCTAGGGATGGACGG 3980  
QY 61 GATTTCATGATCAGTATCCGATTGTCACAAAGAAACGATTGATCGAGCGATCAGGTGCTC 120  
DB 3981 GATTTCATGATCAGTATCCGATTGTTAAAGAAACGATTGATCGAGCGATCAGGTGCTA 4040  
QY 121 GGTATGATTACGTTATCTATCGATACGGAAGAGACAACTCAATCAGACCCGCTAT 180  
DB 4041 GGTATGATTGCGTTATCTATCGATACGGAAGAGACAACTCAATCAGACCCGCTAT 4100  
QY 181 ACGCAACAGCGATCTAGCGACTTCGGTTGCTATCTACCGTTTATTGCAAGAAAGGCG 240  
DB 4101 ACGCAACAGCGATCTAGCGACTTCGGTTGCTATCTACCGTTTATTGCAAGAAAGGCG 4160  
QY 241 TATCAGCGTATATGGTTGCTGTTCTCTCTGGAGATPACTCTGCTTGGTGGCAAGC 300  
DB 4161 TATCAGCGTATATGGTTGCTGTTCTCTCTGGAGATPACTCTGCTTGGTGGCAAGC 4220  
QY 301 GCGCCTTGATTTGAAGATCGGTTGCTTGGTCAAGGTGCCAGTATCTAGGATGGACGG 360  
DB 4221 GCGCCTTGATTTGAAGATCGGTTGCTTGGTCAAGGTGCCAGTATCTAGGATGGACGG 4280  
QY 361 GAAGCGCTCCTGCTGACTCTGGCAAGATGGTACGATTTCTCAATACGCCAGTAGAGTTC 420  
DB 4281 GAAGCGCTCCTGCTGACTCTGGCAAGATGGTACGATTTCTCAATACGCCAGTAGAGTTC 4340  
QY 421 ATTGAAGAGCGCTGCAAAAGCTTCTGAACCTGGAGTGGTTACTCCAGCCACTATAAC 480  
DB 4341 ATTGAAGAGCGCTGCAAAAGCTTCTGAACCTGGAGTGGTTACTCCAGCCACTATAAC 4400  
QY 481 ACACCTGCACAAATCGTCAITGCTGGAGAGTGGTTGCACTTGATCGAGCGGTGAACTT 540  
DB 4401 ACACCTGCACAAATCGTCAITGCTGGAGAGTGGTTGCACTTGATCGAGCGGTGAACTT 4460  
QY 541 TTGCAAGAGCAGTGCACAAAGCTTGATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCAACC 600  
DB 4461 TTGCAAGAGCAGTGCACAAAGCTTGATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCAACC 4520  
QY 601 GCTCTCCTTGAGCCTGCTAGCCAGAACTAGCTGAACTCTAGCTCAGGTGAAGTTTCA 660  
DB 4521 TCTCTCCTTGAGCCTGCTAGCCAGAACTAGCTGAACTCTAGCTCAGGTGAAGTTTCA 4580

QY 661 GATTTTACTTGTCCCTAGTGGCAATACAGAGCTGCTGTGATGCACAAAGAGACATT 720  
DB 4581 GATTTTACTTGTCCCTAGTGGCAATACAGAGCTGCTGTGATGCACAAAGAGACATT 4640  
QY 721 GCTCAGCTCTTGACCGCTCAGGTCAAGGAAACCCGTTCCGTTCTCTATGAAAGTATTGGGTC 780  
DB 4641 GCTCAGCTCTTGACCGCTCAGGTCAAGGAAACCCGTTCCGTTCTCTATGAAAGTATTGGGTC 4700  
QY 781 ATGCAAGAGCAGGATTAAGCACTTTATCGAGATTGACCGGGGAAAGTCTTGTCAAGT 840  
DB 4701 ATGCAAGAGCAGGATTAAGCACTTTATCGAGATTGACCGGGGAAAGTCTTGTCAAGT 4760  
QY 841 TTTGTTAAAAAATGATCAAACTGCTCAGTCTAGTCTCATGTTGGAAGATCAAGCGAGTTTA 900  
DB 4761 TTTGTTAAAAAATGATCAAACTGCTCAGTCTAGTCTCATGTTGGAAGATCAAGCGAGTTTA 4820  
QY 901 GTAGCACTTTTAGAAAAATAG 921  
DB 4821 GTAGCACTTTTAGAAAAATAG 4841  
RESULT 7  
LOCUS AX194275 945 bp DNA linear PAT 15-AUG-2001  
DEFINITION Sequence 256 from Patent WO0149721.  
ACCESSION AX194275  
VERSION AX194275.1 GI:15211757  
KEYWORDS Streptococcus pneumoniae.  
SOURCE Streptococcus pneumoniae.  
ORGANISM Streptococcus pneumoniae.  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.  
REFERENCE 1 (bases 1 to 945)  
AUTHORS Dougherty,T.J., Pucci,M.J., Dougherty,B.A., Davison,D.B.,  
Brucoleri,R.E. and Thanassi,J.A.  
TITLE Novel bacterial genes and proteins that are essential for cell  
viability and their uses  
JOURNAL Patent: WO 0149721-A 256 12-JUL-2001;  
Bristol-Myers Squibb Co. (US)  
FEATURES  
Location/Qualifiers  
1..945  
/organism="Streptococcus pneumoniae"  
/db\_xref="taxon:1313"  
BASE COUNT 254 a 205 c 232 g 254 t  
ORIGIN  
Query Match 98.6%; Score 908.4; DB 6; Length 945;  
Best Local Similarity 99.3%; Pred. No. 2.9e-238;  
Matches 912; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 1 ATGACTAAACAGCGCTTTTATTGCTGGTCAAGGTGCCAGTATCTAGGATGGACGG 60  
DB 1 ATGACTAAACAGCGCTTTTATTGCTGGTCAAGGTGCCAGTATCTAGGATGGACGG 60  
QY 61 GATTTCTATGATCAGTATCCGATTGTCACAAAGAAACGATTGATCGAGCGATCAGGTGCTC 120  
DB 61 GATTTCTATGATCAGTATCCGATTGTTAAAGAAACGATTGATCGAGCGATCAGGTGCTA 120  
QY 121 GGTATGATTTAGTTATCTCATCGATACGGAAGAGACAACTCAATCAGACCCGCTAT 180  
DB 121 GGTATGATTTAGTTATCTCATCGATACGGAAGAGACAACTCAATCAGACCCGCTAT 180  
QY 181 ACGCAACAGCGATCTAGCGACTTCGGTTGCTATCTACCGTTTATTGCAAGAAAGGCG 240  
DB 181 ACGCAACAGCGATCTAGCGACTTCGGTTGCTATCTACCGTTTATTGCAAGAAAGGCG 240  
QY 241 TATCAGCGTATATGGTTGCTGTTTGTCTCTTGGAGAAATCTCTGCTTGGTGGCAAGC 300  
DB 241 TATCAGCGTATATGGTTGCTGTTTGTCTCTTGGAGAAATCTCTGCTTGGTGGCAAGC 300  
QY 301 GCGCCTTGATTTGAAGATCGGTTGCTTGGTCAAGGTGCCAGTATCTAGGATGGACGG 360  
DB 301 GCGCCTTGATTTGAAGATCGGTTGCTTGGTCAAGGTGCCAGTATCTAGGATGGACGG 360

QY 361 GAAGCGGCTCTGCTGACTCTCGCAAGATGGTAGCAGTTCTCAATACCCAGTAGAGGTC 420  
 DB 361 GAAGCGGCTCTGCTGACTCTCGCAAGATGGTAGCAGTTCTCAATACCCAGTAGAGGTC 420  
 QY 421 ATTGAAGAAGCGCTGTCAAAAAGCTTCTGAACCTTTGAGTGGTTACTCCAGGCCAATATAAC 480  
 DB 421 ATTGAAGAAGCGCTGTCAAAAAGCTTCTGAACCTTTGAGTGGTTACTCCAGGCCAATATAAC 480  
 QY 481 ACACCTGCACAAATCGTCAATCTCTCGAAGAGTGGTTGCAAGTGTGATCGAGCGGTTGAACCTT 540  
 DB 481 ACACCTGCACAAATCGTCAATCTCTCGAAGAGTGGTTGCAAGTGTGATCGAGCGGTTGAACCTT 540  
 QY 541 TTGCAAGAAGCAGGTCGCAAAAGCTTGTATCTCTTTAAGGTGTGAGGTCCTTTCACACC 600  
 DB 541 TTGCAAGAAGCAGGTCGCAAAAGCTTGTATCTCTTTAAGGTGTGAGGTCCTTTCACACC 600  
 QY 601 GCTCTCCTTTGAGCCTGTAGCCAGAACTAGCTGAAACTCTAGCTCAGTAAAGTTTTCAC 660  
 DB 601 TCTCTCCTTTGAGCCTGTAGCCAGAACTAGCTGAAACTCTAGCTCAGTAAAGTTTTCAC 660  
 QY 661 GATTTTACTTGTCCCTAGTGGCAATACAGAGCTGTGTGATGCAAAAGAGGACATT 720  
 DB 661 GATTTTACTTGTCCCTAGTGGCAATACAGAGCTGTGTGATGCAAAAGAGGACATT 720  
 QY 721 GCTCAGCTCTTGACGCGTCAGCTCAAGAAACCCGTTCTCTTATGAAAGTATTGGGTC 780  
 DB 721 GCTCAGCTCTTGACGCGTCAGCTCAAGAAACCCGTTCTCTTATGAAAGTATTGGGTC 780  
 QY 781 ATGCAAGAAGCAGGATAGCAACTTTATCGAGATTGACCGGGGAAAGTCTTGTCAAGT 840  
 DB 781 ATGCAAGAAGCAGGATAGCAACTTTATCGAGATTGACCGGGGAAAGTCTTGTCAAGT 840  
 QY 841 TTTGTTAAAAAATTGATCAAACTGCTCATTGCTCATTAGCTCATGTGGAAGATCAAGCGATT 900  
 DB 841 TTTGTTAAAAAATTGATCAAACTGCTCATTGCTCATTAGCTCATGTGGAAGATCAAGCGATT 900  
 QY 901 GTAGCATTTTAGAAAA 918  
 DB 901 GTAGCATTTTAGAAAA 918

RESULT 8  
 LOCUS AR193756 1196 bp DNA linear PAT 20-APR-2002  
 DEFINITION Sequence 8 from patent US 6348328.  
 ACCESSION AR193756  
 VERSION AR193756.1 GI:20240348  
 KEYWORDS Unknown.  
 SOURCE Unknown.  
 ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 1196)  
 AUTHORS Black, M. Terence., Hodgson, J. Edward., Knowles, D. Justin, Charles., Nicholas, R. Oakley, and Stodola, R. King.  
 TITLE Compounds  
 JOURNAL Patent: US 6348328-A 8 19-FEB-2002;  
 FEATURES Location/Qualifiers  
 BASE COUNT 355 a 223 c 286 g 332 t  
 ORIGIN 1..1196 /organism="unknown"  
 Query Match 56.8%; Score 523.2; DB 6; Length 1196;  
 Best Local Similarity 99.1%; Pred. No. 1.6e-132;  
 Matches 547; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 372 TGTGACTCTGCAAGATGGTAGAGTTCTCAATACGCCAGTAGAGGTCATTGAAGAAGC 431  
 DB 1 TGTGAACTTGGCAAGATGGTAGAGTTCTCAATACGCCAGTAGAGGTCATTGAAGAAGC 60  
 QY 432 CTGTCAAAAAGCTTCT-GAAGTTGAGTGGTTACTCCAGCCACTATTAACACCTGCAC 490

DB 61 CTGTCAAAAAGCTTCTGGAACCTTGGAGTGGTTACTCCAGCCCACTATAACACACCTGCAC 120  
 QY 491 AAATCGTCAATGCTGGAGAAGTGGTTGCAAGTTGATCGACGGTGAACCTTTTGCAGAAG 550  
 DB 121 AAATCGTCAATGCTGGAGAAGTGGTTGCAAGTTGATCGACGGTGAACCTTTTGCAGAAG 180  
 QY 551 CAGGTCCCAAGCGCTTGAATCTCTTAAGGTGTGAGGTCCTTTCACACCGCTCTCCTTG 610  
 DB 181 CAGGTCCCAAGCGCTTGAATCTCTTAAGGTGTGAGGTCCTTTCACACCGCTCTCCTTG 240  
 QY 611 AGCTCTGCTAGCCAGAACTAGCTGAAACTCTAGCTCAGGTAAGTTTTCAGATTTTACTT 570  
 DB 241 AGCTCTGCTAGCCAGAACTAGCTGAAACTCTAGCTCAGGTAAGTTTTCAGATTTTACTT 300  
 QY 671 GTCCCTAGTCCGCAATACAGAGCTGTGTGATGCAAAAGAGGACATTGCTCAGCTCT 730  
 DB 301 GTCCCTAGTCCGCAATACAGAGCTGTGTGATGCAAAAGAGGACATTGCTCAGCTCT 360  
 QY 731 TGACGGCTCAGGTCAGGAAACCCGTTCTGTTTCTATGAAAGTATTGGGGTCATGCAAGAAG 790  
 DB 361 TGACGGCTCAGGTCAGGAAACCCGTTCTGTTTCTATGAAAGTATTGGGGTCATGCAAGAAG 420  
 QY 791 CAGGCATAGCAACTTTA-TCGAGATTGACCGGGGAAAGTCTTGTCAAGTTTGTATAA 849  
 DB 421 CAGGCATAGCAACTTTA-TCGAGATTGACCGGGGAAAGTCTTGTCAAGTTTGTATAA 480  
 QY 850 AAAATTGATCAAACTGCTCATTAGCTCATGTGGAAGATCAAGCGAGTTTAGTAGCACTT 909  
 DB 481 AAAATTGATCAAACTGCTCATTAGCTCATGTGGAAGATCAAGCGAGTTTAGTAGCACTT 540  
 QY 910 TTAGAAAAATAG 921  
 DB 541 TTAGAAAAATAG 552

RESULT 9  
 LOCUS AE006603/c 10909 bp DNA linear BCT 01-JUN-2001  
 DEFINITION Streptococcus pyogenes M1 GAS strain SF370, section 132 of 167 of the complete genome.  
 ACCESSION AE006603 AE004092  
 VERSION AE006603.1 GI:13622797  
 KEYWORDS Streptococcus pyogenes M1 GAS.  
 SOURCE Streptococcus pyogenes M1 GAS.  
 ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.  
 REFERENCE 1 (bases 1 to 10909)  
 AUTHORS Ferretti, J.J., McShan, W.M., Adjic, D., Savic, D., Savic, G., Lyon, K., Primeaux, C., Sezate, S., Surorov, A.N., Kenton, S., Lai, H., Lin, S., Qian, Y., Jia, H.G., Najjar, F.Z., Ren, Q., Zhu, H., Song, L., White, J., Yuan, X., Clifton, S.W., Roe, B.A., and McLaughlin, R.E.  
 TITLE Complete genome sequence of an M1 strain of Streptococcus pyogenes  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (8), 4658-4663 (2001)  
 MEDLINE 21192684  
 PUBMED 11296296  
 REFERENCE 2 (bases 1 to 10909)  
 AUTHORS Ferretti, J.J., McShan, W.M., Adjic, D., Savic, D., Savic, G., Lyon, K., Primeaux, C., Sezate, S., Surorov, A.N., Kenton, S., Lai, H., Lin, S., Qian, Y., Jia, H.G., Najjar, F.Z., Ren, Q., Zhu, H., Song, L., White, J., Yuan, X., Clifton, S.W., Roe, B.A., and McLaughlin, R.E.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-APR-2001) Department of Microbiology and Immunology, University of Oklahoma Health Sciences Center, 940 SL Young Blvd, Oklahoma City, OK 73104 USA  
 FEATURES Location/Qualifiers  
 source 1..10909  
 /organism="Streptococcus pyogenes M1 GAS"  
 /strain="SF370"  
 /serotype="M1"  
 /db\_xref="taxon:160490"  
 complement(142..912)  
 /gene="acca"

```

CDS
/note="SPY1743"
/complement(142..912)
/gene="accA"
/EC_number="6.4.1.2"
/note="Best Blastp hit = dbj|BAB06884.1| (AP001517)
acetyl-CoA carboxylase carboxyltransferase alpha subunit
[Bacillus halodurans]"
/codon_start=1
/transl_table=11
/product="putative acetyl-CoA carboxylase alpha subunit"
/protein_id="AAK34487.1"
/db_xref="GI:13622798"
/translacion="MTDVSRLKEARDQGRLLTLDVANLIFDDFMELGHRHFDGGA
IVGLAYLAQCPVTYIGIQKQLQDLNARFQCPNPEGYRKALRLMKOAEKFRFVY
TFNTAGYGVGAERGGGEATAKNLMESDLKVPILAIIGEGSGGALALAVADQ
VMLNFMVAVLSPGPFASILWKDGGRAEAAELMKITAGELYKMGIVDRIIPEHGYF
SSEIVDIKANLIEQITSLQAKPLDQLLDERYQRFKY"
/complement(909..1775)
/gene="accD"
/note="SPY1744"
/complement(909..1775)
/gene="accB"
/note="Best Blastp hit = sp|054776|ACCD SYNPF7
ACTYL-COENZYME A CARBOXYLASE CARBOXYL TRANSFERASE SUBUNIT
BETA (ACCASE BETA CHAIN) >gi|1399855|gb|AAB82034.1|
(US9237) carboxyltransferase beta subunit [Synechococcus
PCC7942]"
/codon_start=1
/transl_table=11
/product="putative acetyl-CoA carboxylase beta subunit"
/protein_id="AAK34488.1"
/db_xref="GI:13622799"
/translacion="MALFRKKDKYIRITPNNSLKGVSNNVPEVDELPAKCPACKHM
IYKDLGLACITPCTSNFRISAQERLLTLDGSEFQELTSETDPLRFPYQFKL
QKAKETGLHEAVLTGMAYKEQKIALIMDSHFIMASGTVVGEKITLPLFLAEEN
LPVYITFASGARMQEGISLMQMAKVSAAVKHNSAGFLYTLITDPTTGGVTASFA
MEGDIILAEQPSQLVGFAGRVETTVRENLPDFOKAELFDQHGFDVAIVKRTLRDK
IAHLVAFHGGGQ"
/complement(1784..3148)
/gene="accC"
/note="SPY1745"
/complement(1784..3148)
/gene="accC"
/note="The N-terminal amino acid sequence of this ORF has
been determined from a spot isolated by 2-D gel
electrophoresis from another strain of S. pyogenes. Hogan,
D. A., Whitton, M. M., Rogers, J. and R. A. VanBogelen.
2000. Two-dimensional gel electrophoresis map of
Streptococcus pyogenes proteins. Unpublished data.; Best
Blastp hit = pir|A69581|acetyl-CoA carboxylase (biotin
carboxylase subunit) accC - Bacillus subtilis
>gi|1303913|dbj|BAAL12569.1| (D84432) ybhX [Bacillus
subtilis] >gi|2634868|emb|CAB14365.1| (Z99116) acetyl-CoA
carboxylase subunit (biotin carboxylase subunit) [Bacillus
subtilis]"
/codon_start=1
/transl_table=11
/product="putative acetyl-CoA carboxylase biotin
carboxylase subunit"
/protein_id="AAK34489.1"
/db_xref="GI:13622800"
/translacion="MFKKILIANRGEIAVRIIRAAARELGISTVAVYSEADKEALHTIL
ADPAICIGPARKSESYLNNSVLSSAIVTGAQAIHPGFGFLSENSKFTATCEEMNKF
IGPSASVMDQMGKINARMEIKAGVPVPGDGVYNAQEAALANKIGVPMLKAS
AGGGGKIRKVEATEALEAFNAASOEGALGFGNGAMYLEKIVYPARHTEVOLLDAY
GNIIHIGERDCSILORNNQVLRPESPSIAIGNTLRHMGQAARAAENAVENAGTIEF
LLDESEKVFYEMNTRIQUVHPHTVGVDIKVEQIKIAAGQALANQEDITITGH
ALBCRNAENANTFAPSPGKITDLYMPFGSGVGLRVDSAVNGYAIPIPYDSMIAKII
VHGSNFDLMKRRQALVLEIEGIIITDFOLDLISDKRVAGDYDTSFLMETTLPH
YNKD"
/complement(3180..3599)
/gene="fabZ"
/note="SPY1746"

CDS
/note="SPY1743"
/complement(142..912)
/gene="accA"
/EC_number="6.4.1.2"
/note="Best Blastp hit = dbj|BAB06884.1| (AP001517)
acetyl-CoA carboxylase carboxyltransferase alpha subunit
[Bacillus halodurans]"
/codon_start=1
/transl_table=11
/product="putative acetyl-CoA carboxylase alpha subunit"
/protein_id="AAK34487.1"
/db_xref="GI:13622798"
/translacion="MTDVSRLKEARDQGRLLTLDVANLIFDDFMELGHRHFDGGA
IVGLAYLAQCPVTYIGIQKQLQDLNARFQCPNPEGYRKALRLMKOAEKFRFVY
TFNTAGYGVGAERGGGEATAKNLMESDLKVPILAIIGEGSGGALALAVADQ
VMLNFMVAVLSPGPFASILWKDGGRAEAAELMKITAGELYKMGIVDRIIPEHGYF
SSEIVDIKANLIEQITSLQAKPLDQLLDERYQRFKY"
/complement(909..1775)
/gene="accD"
/note="SPY1744"
/complement(909..1775)
/gene="accB"
/note="Best Blastp hit = sp|054776|ACCD SYNPF7
ACTYL-COENZYME A CARBOXYLASE CARBOXYL TRANSFERASE SUBUNIT
BETA (ACCASE BETA CHAIN) >gi|1399855|gb|AAB82034.1|
(US9237) carboxyltransferase beta subunit [Synechococcus
PCC7942]"
/codon_start=1
/transl_table=11
/product="putative acetyl-CoA carboxylase beta subunit"
/protein_id="AAK34488.1"
/db_xref="GI:13622799"
/translacion="MALFRKKDKYIRITPNNSLKGVSNNVPEVDELPAKCPACKHM
IYKDLGLACITPCTSNFRISAQERLLTLDGSEFQELTSETDPLRFPYQFKL
QKAKETGLHEAVLTGMAYKEQKIALIMDSHFIMASGTVVGEKITLPLFLAEEN
LPVYITFASGARMQEGISLMQMAKVSAAVKHNSAGFLYTLITDPTTGGVTASFA
MEGDIILAEQPSQLVGFAGRVETTVRENLPDFOKAELFDQHGFDVAIVKRTLRDK
IAHLVAFHGGGQ"
/complement(1784..3148)
/gene="accC"
/note="SPY1745"
/complement(1784..3148)
/gene="accC"
/note="The N-terminal amino acid sequence of this ORF has
been determined from a spot isolated by 2-D gel
electrophoresis from another strain of S. pyogenes. Hogan,
D. A., Whitton, M. M., Rogers, J. and R. A. VanBogelen.
2000. Two-dimensional gel electrophoresis map of
Streptococcus pyogenes proteins. Unpublished data.; Best
Blastp hit = pir|A69581|acetyl-CoA carboxylase (biotin
carboxylase subunit) accC - Bacillus subtilis
>gi|1303913|dbj|BAAL12569.1| (D84432) ybhX [Bacillus
subtilis] >gi|2634868|emb|CAB14365.1| (Z99116) acetyl-CoA
carboxylase subunit (biotin carboxylase subunit) [Bacillus
subtilis]"
/codon_start=1
/transl_table=11
/product="putative acetyl-CoA carboxylase biotin
carboxylase subunit"
/protein_id="AAK34489.1"
/db_xref="GI:13622800"
/translacion="MFKKILIANRGEIAVRIIRAAARELGISTVAVYSEADKEALHTIL
ADPAICIGPARKSESYLNNSVLSSAIVTGAQAIHPGFGFLSENSKFTATCEEMNKF
IGPSASVMDQMGKINARMEIKAGVPVPGDGVYNAQEAALANKIGVPMLKAS
AGGGGKIRKVEATEALEAFNAASOEGALGFGNGAMYLEKIVYPARHTEVOLLDAY
GNIIHIGERDCSILORNNQVLRPESPSIAIGNTLRHMGQAARAAENAVENAGTIEF
LLDESEKVFYEMNTRIQUVHPHTVGVDIKVEQIKIAAGQALANQEDITITGH
ALBCRNAENANTFAPSPGKITDLYMPFGSGVGLRVDSAVNGYAIPIPYDSMIAKII
VHGSNFDLMKRRQALVLEIEGIIITDFOLDLISDKRVAGDYDTSFLMETTLPH
YNKD"
/complement(3180..3599)
/gene="fabZ"
/note="SPY1746"

CDS
/note="Best Blastp hit = dbj|BAB16042.1| (AB030883)
similar to Bacillus subtilis hydroxymyristoyl-(acyl
carrier protein) dehydratase [Streptococcus
zooepidemicus]"
/codon_start=1
/transl_table=11
/product="putative beta-hydroxyacyl-ACP dehydratase"
/protein_id="AAK34490.1"
/db_xref="GI:13622801"
/translacion="MDIREIQALPHRYPMLLVDRLVLESDDHIVAIAKNVTINEPFFN
GHFPHYPMKQVLTMEALQTAGVLELSKEENKGLVYFAGMDKVKFKKQVVPDQLV
MTAFIPKRGITIAVVEAAEAVDGLAASGTTUTPACGQ"
/complement(3599..4099)
/gene="accB"
/note="SPY1747"
/complement(3599..4099)
/gene="accB"
/note="Best Blastp hit = pir||C75558 acetyl-CoA
carboxylase, biotin carboxyl carrier protein - Deinococcus
radiodurans (strain R1)
>gi|6457783|gb|AAF09708.1|AE001874.5 (AE001874) acetyl-CoA
carboxylase, biotin carboxyl carrier protein [Deinococcus
radiodurans]"
/codon_start=1
/transl_table=11
/product="putative biotin carboxyl carrier protein"
/protein_id="AAK34491.1"
/db_xref="GI:13622802"
/translacion="MNIQEIYKDLMAQFDTSLSRLFKTNKTEGELIFSKNEOHLNASTS
NQEHAVPVPQVLPNTPASBASPASVSKDVVEEQPAQSFVAAGSDIVESPLVGVAY
LAAQSPDPFPVAVGDTVKKGQTLVIIEMAKMNEVPAPCDGVTEILVSNEDVIEFGQ
GLVRIK"
/complement(4101..5333)
/gene="fabF"
/note="SPY1748"
/complement(4101..5333)
/gene="fabF"
/EC_number="2.3.1.41"
/note="The N-terminal amino acid sequence of this ORF has
been determined from a spot isolated by 2-D gel
electrophoresis from another strain of S. pyogenes. Hogan,
D. A., Whitton, M. M., Rogers, J. and R. A. VanBogelen.
2000. Two-dimensional gel electrophoresis map of
Streptococcus pyogenes proteins. Unpublished data.; Best
Blastp hit = pir||G69842 3-oxoacyl-(acyl-carrier-protein)
synthase [EC 2.3.1.41] - Bacillus subtilis
>gi|2633471|emb|CAB12975.1| (Z99109) similar to 3-oxoacyl-
acyl-carrier protein synthase [Bacillus subtilis]
>gi|2633488|emb|CAB12991.1| (Z99110) similar to 3-oxoacyl-
acyl-carrier protein synthase [Bacillus subtilis]"
/codon_start=1
/transl_table=11
/product="putative beta-ketoacyl-ACP synthase II"
/protein_id="AAK34492.1"
/db_xref="GI:13622803"
/translacion="MTEKRVVVTGYGLTSPIGHDPETFWNNLKAGQIGIPITKFDTT
DIAVNAAEICDPFFDKYFKKQDLNFRDISGLYALAAKBAINHADINIMVSDRFRG
VIVASGIGGIAIEIEQVIRLHEKQPKRVKPTLPKALPNMAAGNVAMSLKAQGVCKSI
NTACASNDIAGDAFRAIKFTGQDVMIVGSGEAAITKFAIQSUTALSTTDPSSR
SIPDKDRNGFIMGESGMLVLESLHAQERGATILAEIVGYGNTCDAYHMTSPNPEG
LGAKATHALQAEIASEASAINVNAHGTSTPANKEGSOAIVAVLGKVDVPSKSF
TGHLGAAGAIEATATIEAVRNHVVPMTAGTQALSEIDIEANVIFGEGKETAINVAISN
TFGFGHNAVLAKMEE"
/complement(5348..6082)
/gene="fabG"
/note="SPY1749"
/complement(5348..6082)
/gene="fabG"
/EC_number="1.1.1.100"
/note="The N-terminal amino acid sequence of this ORF has
been determined from a spot isolated by 2-D gel

```

electrophoresis from another strain of *S. pyogenes*. Hogan, D. A., Whittom, M. M., Rogers, J. and R. A. VanBogelen. 2000. Two-dimensional gel electrophoresis map of *Streptococcus pyogenes* proteins. Unpublished data.; Best Blastp hit = dbj|BAB06210.1| (AP001515)

3-oxoacyl-(acyl-carrier protein) reductase [Bacillus halodurans]  
/codon\_start=1  
/transl\_table=11  
/product="putative beta-ketoacyl-ACP reductase"  
/protein\_id="AAK34493.1"  
/db\_xref="GI:13622804"  
translation="WEIKGKNIFFITGSTRIGLAWHOFASLEANIVLNGRSALSSEL  
VASFTDYGVTVTISDYSSEASEAKWVNEALSISDVLVNNAGITNDKMLKWE  
EDPERVLKINTGARNMTQSVLKPIMKARQCALINVSIVGLTGNIGQANYASAKAGM  
IGFTKSAREVAARNICVNAIPFISDMTGVLPFKMQEQILSIQIPMKRIGKAQEYA  
HLASFLVQDYITGVIAIDGGMNQ"  
complement(16072..7010)

gene

Query Match 47.7%; Score 439.4; DB 1; Length 10909;  
Best Local Similarity 67.3%; Pred. No. 1.8e-109;  
Matches 620; Conservative 0; Mismatches 301; Indels 0; Gaps 0;

QY 1 ATGACTAAACAGCGCTTTTATTTCTGCTGCTCAAGTGCCTCAGTATCTAGGATGGACGG 60  
DB 7010 ATGACAAAGACAGCGCTTTTATTTCTGCTGCTCAAGTGCCTCAGTATCTAGGATGGACGG 6951

QY 61 GATTCTTATGATCAGTATCGGATTCCTCAAGAAAGATTGATCGAGCGAGTCAGTGTCTC 120  
DB 6950 GATTCTTATGATCAGTATTCCTATTCAGAAAGATTGATCGAGCGAGTCAGTGTCTC 6891

QY 121 GATTATGATTTACGTTATCTATCATCGATACGGAAGAACAACTCAATCAGACCGGCTAT 180  
DB 6890 GGATACGATTTTCGCGCGTTTGTATTCATAGTACGAGTTTAAACTTAATCAAACTAGCTAT 6831

QY 181 ACGCAACAGCATCTAGCGACTTCGGTGTCTATCTACCGTTTATTCGACAGAAAGGCG 240  
DB 6830 ACGCAACAGCATTTTGTGATCATCAATTCGTTATTTACCGTGTTCGCGCTTACATCAC 6771

QY 241 TATCAGCTGATATGTTGCTGCTGTTGTTCTCTTGGAGAAATCTCTGCTTGGTGGCAGC 300  
DB 6770 GTTAAACCGGATGATGATGCTGCTTCTTCTTAGGAGAAATCTCAGCTTTGGTAGCATCA 6711

QY 301 GCGCGCTTGGATTTTGTAGATCGGTTGCTTGTAGCTAGCTAGCGTGGAGCTTATGGAA 360  
DB 6710 GGGGCACTCTCTTTTGAAGATACCTCTATCTCTAGTAGCTAGAGAGCGCGCTTATGGAG 6651

QY 361 GAAGCGGCTCTGCTGACTCTCGCAAGATGGTAGAGTTCTCAATACGCCAGTAGAGTCTC 420  
DB 6650 GAAGCAGCACCAAGGATCTGGAAGATGGTTGCCGTTATGATACAGAGCTGCAAGTC 6591

QY 421 ATGGAAGAGCTGTCAAAAGCTTCTGAACTTGGAGTGGTTTACTCCAGGCCAATATAAC 480  
DB 6590 ATCGAAGAGTGTCTCAAAATAGCTCTGCTTAAACATGAGTGGTTGCTTCCAGCAAACTATAAT 6531

QY 481 ACACCTGCAAAATCGTCAATCTCTCGAAGTGGTTGAGTTGATCGAGCGGTTGAACTT 540  
DB 6530 ATCTCTAGTCAAAATGTAATTTGGTGTGACAGAGATGCTGTGAACGTGGCAGTTGAACCTT 6471

QY 541 TTGCAAGAGAGGTCGCAAAAGCTTGAATCTCTTAAAGGTGTCAAGGTCCTTTTCAACC 600  
DB 6470 TTAAGGAAAGGGAGTAAAGGTTTAAATCCCTTTTAAACGTGTCAAGGTCCTTTTCAACT 6411

QY 601 GCTCTCTTGGCTGTAGCGGAGTACAGCAAGATAGCTGAACTCTAGCTCAGTAAAGTTTCA 660  
DB 6410 GTTTTGTAGAACAGTACAGTACGCTTGTGGTAAAGAGTGGAAAGATACAATCTCAGT 6351

QY 661 GATTTTACTTGTCCCTAGTCGGCAATACAGAGCTGTGTGTGATGCAAAAGAGGACATT 720  
DB 6350 GACTTCAAGATTCCTTTGGTGTGTAATACCGAAGCTAATATTATGGAAGAAAGACCGTATC 6291

QY 721 GCTCAGCTCTTACGGGTCAGTCAAGAAACCGTTCGTTCTCTATGAAGTATTTGGGTC 780

DB 6290 CAGAACTATTAGCCCGTCAAGTCAATGAGAGCCTGTTGTTTTATGACAGTGTGGACT 6231

QY 781 ATGCAAGAAGCAGGCAATAACAACTTTATCGAGATTGGACCGGGAAGCTTTGTCAAGT 840

DB 6230 TTAGTAGAAGTGGCAATAACAACTTTATCGAGTAGGACCAAGTAAAGTTTGTACAGT 6171

QY 841 TTGTTTAAAAAATTCATCAAACTGCTCATTCTAGCTCATGTGGAAGATCAACGAGTTTA 900

DB 6170 TTGTGAAGAAATTCATCAAACTTACTATGTTAGTGTGTGAAGAAATGTTAAGCTTA 6111

QY 901 GTAGCACTTTTAAAAAATAG 921

DB 6110 CGTTATTTTATAGATAGAG 6090

RESULT 10  
AE014165/c  
LOCUS  
DEFINITION  
50416 bp DNA linear BCT 19-JUL-2002  
Streptococcus pyogenes MGAS315, section 30 of 37 of the complete genome.  
ACCESSION  
AE014165 AB014074  
VERSION  
AE014165.1 GI:21905237  
KEYWORDS  
SOURCE  
ORGANISM  
Streptococcus pyogenes MGAS315.  
Streptococcus pyogenes MGAS315  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.  
REFERENCE  
1 (bases 1 to 50416)  
Beres, S.B., Sylva, G.L., Barbican, K.D., Lei, B., Hoff, J.S., Mammarella, N.D., Liu, M.-Y., Smoot, J.C., Porcella, S.F., Parkins, L.D., McCormick, J.K., Leung, D.Y.M., Schlievert, P.M. and Musser, J.M.  
Genome sequence of a serotype M3 strain of group A *Streptococcus*: phage-encoded toxins, the high-virulence phenotype, and clone emergence  
Proc. Natl. Acad. Sci. U.S.A. 99 (15), 10078-10083 (2002)  
JOURNAL  
REFERENCE  
2 (bases 1 to 50416)  
Beres, S.B., Sylva, G.L., Barbican, K.D., Lei, B., Hoff, J.S., Mammarella, N.D., Liu, M.-Y., Smoot, J.C., Porcella, S.F., Parkins, L.D., McCormick, J.K., Leung, D.Y.M., Schlievert, P.M. and Musser, J.M.  
Direct Submission  
Submitted (14-JUN-2002) Laboratory of Human-Bacterial Pathogenesis, Rocky Mountain Laboratories, NIAID, NIH, 903 South Fourth St., Hamilton, MT 59840, USA  
FEATURES  
Location/Qualifiers  
source  
1..50416  
/organism="Streptococcus pyogenes MGAS315"  
/strain="MGAS315"  
/serotype="M3"  
/db\_xref="taxon:198466"  
/note="group: A"  
117..1103  
/gene="Spym3\_1493"  
117..1103  
/gene="Spym3\_1493"  
/note="best Elastp hit: gb|BAJ98256.1| (AB010083) putative esterase [Streptococcus pyogenes MGAS315]"  
/codon\_start=1  
/transl\_table=11  
/product="putative esterase"  
/protein\_id="AAM80100.1"  
/db\_xref="GI:21905238"  
/translation="MKHPRTKTLVTLGLLLTLCLPTAAASRSKSWFIEQYFWLKR  
DKSYKQDDPSFQRYLDACREQSKPYQLDNLVGLVQENLYCMQVYWNDRKPK  
LQYTHYKQNVFLMGDSAGGALGLAHNESPQPKQLVLSPLMDVMTSHPEI  
PYREDADPTLSSGLKRVGLWAYSDNTNHYVSPKNGPITLPLTFTGTREIFY  
PIQRYAAKLKAENHITFIQEGNHHVPIPIEAKTAQYQIIDAINKTP"  
complement(1239..1589)  
/gene="rbfa"  
/note="Spym3\_1494"  
complement(1239..1589)  
CDS  
gene  
CDS

```

/note="rbfa"
/note="best blastp hit: gb|AAK34465.1| (AE006601) putative
ribosome binding factor A [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative ribosome binding factor A"
/protein_id="AAK80101.1"
/db_xref="GI:21905239"
/translation="MANHRIIDRVGMEIKREVNDILOKVDPDPOVQVITTEVOMOGDL
SLAKVYTIISDLASDNQKQATGLEKATGIKRELKQLTMYKIPDLVFEKNSIAYG
NKIDQLLRDLNKG"
/complement(1795..4599)
/note="infB"
/note="infB"
/complement(1795..4599)
/note="infB"
/note="infB"
/complement(1795..4599)
/note="best blastp hit: gb|AAK34466.1| (AE006601) putative
initiation factor 2 [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative initiation factor 2"
/protein_id="AAK80102.1"
/db_xref="GI:21905240"
/translation="MVEHAKYLGIDVKSASHASVEADAKIISFASKAPDVTASQT
VKPEVAQPSVVVVKETGSEHVEKTVQSKPSKRNFKAEAREAKAQARKQAGSSHR
SQERGGYQPNNHQTEQDGRITHRSGDTNDKIERKASVSPRHQHLQVGDNR
RSFAKENHKGRTNQKQGPQSKPQIDFKAAALKAALKAQNAEYRQSETRFA
QOEAKRLAELARQAEALKAQAEEMSHREAAKSTEEAEATKLKSNSTSAKSTADNR
RKQAKPEKRELTHSGEGKKNKSWNSQNVQKNSWNKNTKKGKNAKNTN
TAPKTVRFKPELPEFEYTEGTUVAEIAKRIKREPAELVKLFPMGWYATONSLD
GDTIELMDVYGEAKKVEVDADIERFFEDENYNPENIVERAPVITMGHVRGK
TTLTLNRNRSVATGEAGGITHGAYQIEEAGKKTFLDPGHAAFTSWRAGASVT
DITILVAADGVMPTIEAINSKAAGVPIIVAINIKDKPANGPERVIAEAGSYII
STAWGCEFEVLSAKENIDDELLETVLLVAEVELKADPTVRAIGTVIEARLDKGG
GAIALVLLQOGLTHVQPIVVGNTFGFRVAMNDLGRVKASAPSTPVTIGLNETPM
ADGFVAYVAAEAKAAGERSKALLKQNTQVSLDNLFDTLKAGETKTVNVIK
ADVGSVEALAAAKVTEVSGRVNVVHSAVGAINSDDVTLLAESNAVLIGENVPTP
QARQATDDVERLHSLIIYKVIIEVEEAMKGLDPVYQKILGEALIIIRETFKVKVG
TIGGFWMVINGKVRDSSVRVIRDSVIFDGLKSLAKHYKDDVKEVNAQEGGLMIENF
NDLKVDITIEAYINEEIVRK"
/complement(4676..4978)
/note="SpvM3_1496"
/note="SpvM3_1496"
/complement(4576..4978)
/note="best blastp hit: gb|AAL98259.1| (AE010083) putative
ribosomal protein [Streptococcus pyogenes MGAS8232]"
/codon_start=1
/transl_table=11
/product="putative ribosomal protein"
/protein_id="AAK80103.1"
/db_xref="GI:21905241"
/translation="MTNLERLSLIGLAQKAGKVISGEELVVKAIHQHQVILVFLAND
AGPNVTKKVIDKSNYINVESTVLNLELSAALGPKRKAATADAGFSKMRITLME"
/complement(4971..5267)
/note="SpvM3_1497"
/note="SpvM3_1497"
/complement(4971..5267)
/note="best non-GAS blastp hit: gb|AAK74710.1| (AE007365)
conserved hypothetical protein [Streptococcus pneumoniae
TIGR4], and gb|AAK99283.1| (AE008427) Conserved
hypothetical protein [Streptococcus pneumoniae R6]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK80104.1"
/db_xref="GI:21905242"
/translation="MSKVKKIPLRKSLVSGEIIAKRDLIRIVTKDQGVFIDPTGKQK
GRGAYIKLDNQEALMAKKKQVFNRSFMDIPESFYDDLTAAYVDHKIKRRELGLD"
/complement(5283..5428)
/note="nuaA"
/note="nuaA"
/complement(5283..5428)
/note="best blastp hit: gb|AAK34469.1| (AE006601) putative
transcription termination-antitermination factor
[Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative transcription
termination-antitermination factor"
/protein_id="AAK80105.1"
/db_xref="GI:21905243"
/translation="MLEAFRILEEKHDKADIIDAVTIESLSKAYKRYRGQSECVIE
FNEKPADQVTVREVEVEEVPDSLEISLKDALAISAYELGDKIRFBESVNEFRVA
AQSAKQITMEKVRQREVMFENYKEHGEIMTGTVERFDQRFIYNIGLSLEAQLSHQ
DOIIGETPKSHRIDVYVYKVENNPKGVNFSRSHPEFKRIIMEQETPEYFDGTEIPI
MSVREAGDRTKVAVRSHNPVNDAGITIVGRGGSNIKVKISKPHKPKRVDATKIGLEIPV
BENIDVIQWDDPAEFIYNAJAPAEVMDLFDDEDLKRAVTVPVDSKLSLAIGRRGQN
VRLAHLFGYRIDIKSASEYDRLEAKAEAAATAVEPVVDDAMTAEEV"
/complement(6615..7151)
/note="SpvM3_1499"
/note="SpvM3_1499"
/complement(6615..7151)
/note="best non-GAS blastp hit: gb|AAK99281.1| (AE008427)
Conserved hypothetical protein [Streptococcus pneumoniae
R6]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK80106.1"
/db_xref="GI:21905244"
/translation="WDSOGPILEKSIKIEBVIKANTSIIDIVTKTVTPRIKAPVEL
VDVEYDKMSGYILSLVDKEDGITVEPDSLTNLIISPLDTIDPDPNPNOYMLEVSS
PGLERPLKTASLKAAGSYINVSILYQALDKVKVFGQLLAFDGETLIDILDKTRHK
IVNIPYQAVAKVRMAVKU"
/complement(7259..7345)
/note="trNA-Ser"
/note="trNA-Ser"
/complement(7410..8045)
/note="SpvM3_1500"
/note="SpvM3_1500"
/complement(7410..8045)
/note="best blastp hit: gb|AAL98263.1| (AE010083) putative
methyltransferase [Streptococcus pyogenes MGAS8232]"
/codon_start=1
/transl_table=11
/product="putative methyltransferase"
/protein_id="AAK80107.1"
/db_xref="GI:21905245"
/translation="MRVRKGAEEHLANNPHYVLINPEDAKGRHHDVFGNDRPHIE
VSGSGKGFITGALKNPNDIYIGIDILQSLVSLYALDKVLASEVNPVKLLRVDGSLTN
YFEDGEVDMYINFDSDPWFKTKERKRLTYKDLTYKRIILPEHGEIHFKTDNRGLFE
YSLASFQYQGMTLRQIWLDLHASNYEGNVMTVEYEEKFSNKQVIYRVEANF"
/complement(8045..8836)
/note="SpvM3_1501"
/note="SpvM3_1501"
/complement(8045..8836)
/note="best blastp hit: gb|AAK99279.1| (AE008427)
Conserved hypothetical protein [Streptococcus pneumoniae
R6]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK80108.1"
/db_xref="GI:21905246"
/translation="MTTTEQELTTLPLRGSKGAYKGYTPNGECVFIKLNTTLPAL
AKEQAPOLLWAKRMGNDMMASQELNGLRTLKEDMNSKQIHLRLHKSRLVNO
LLQNLNYENPYDILLVPEQNALQIQNSYLOAIVKELKSLPEFKSEVATIVHGDII
KHSNWLTITSGMIFLVDVDSVRLTDRMYDVAVLLSHVTPRSRSEWLSYGYKNNDKV
MQKLIWYQPSHLTQLKCFDKRDMEHVQNIYALRKREIFRKK"
/complement(8900..9934)
/note="SpvM3_1502"
/note="SpvM3_1502"
/complement(8900..9934)
/note="best blastp hit: gb|AAK34473.1| (AE006601) putative

```

Query Match

47.7%; Score 439.4; DB 1; Length 50416;

Best Local Similarity 67.3%; Pred. No. 1.9e-109;		Matches 620; Conservative 0; Mismatches 301; Indels 0; Gaps 0;	
QY	1	ATGACTAAACAGCCTTTTATTGCTGCTGAAGTGCCCGATCTCTAGGATGGAGCG	60
Db	30296	ATGACAAGACAGCCTTTTATTGCTGCTGAAGTGCTCAAAAATTAGGATGGCAAG	30237
QY	61	GATTCTTATGATCAGTATCCGATTGTCAAGAAAAGATTGATCGAGCGAGTCAGTGTCT	120
Db	30236	GATTCTTATGATACTTCTGCTATTGTAGAAAAACCTTTGATCAAGCTAGTCAAGTATTG	30177
QY	121	GGTTATGATTAGCTTATCTCATCGATCGGAAGAACAACTCAATCAGACCGCTAT	180
Db	30176	GGATAGATTTCGGCGTTTGTATGATAGTACGAGTTTAAACTTATCAAACTAGCTAT	30117
QY	181	ACGCACACGCAATCTACGACTTCGGTGTCTATCTACCGTTTATTGCAAGAAAGGC	240
Db	30116	ACGCACACGCTATTGTGACATCAATGCTATTACCGTGTTCGGCTTACATCAC	30057
QY	241	TATCAGCTGATATGTTGCTGGTTGCTCTTGAGAAATCTCTGCTTGGTGGCAAGC	300
Db	30056	GTTAAACCGGATATGCTAGTCTGCTTCTCTTAGGAGAAATCTCAGCTTGGTAGCATCA	29597
QY	301	GGCGCTTGATTGAGATCGGTTGCTTGGTAGCTAGCGTGGAGCCTATATGAA	360
Db	29996	GGGCACTCTCTTTGAAGATACCTTATCTCTTAGTAGCTAAGAGAGCGCTTGTATGAG	29937
QY	361	GAAGCGGCTCTGCTGACTCTGGCAGATGTPAGCAGTTCTCAATACGCCAGTAGAGTC	420
Db	29936	GAAGCAGCACCAAGGATCTGGAAGATGTTGCCGTTATGAATACAGAGCTGCAAGTC	29877
QY	421	ATTGAGAGCTGTCAAAAACCTTCTGAACTTGAGTGGTACTCCAGCAACTATAAC	480
Db	29876	ATCGAAGAGTCTGCAATAGTCTGTAACATGAGTGGTGTCTCCAGCAACTATAAT	29817
QY	481	ACACCTGCACAAATGCTCATCTCTCGAAGTGGTTCAGATGTATCGAGCGGTTGAAGTT	540
Db	29816	ACTCTAGTCAATTTGTAATGTTGGTGCAGACAGATGCTGTGACGTGCGAGTTGAAGTT	29757
QY	541	TTGCAAGAGCAGGTGCCAAGCTTTGATCTCTTAAGGTGTCAAGTCCCTTTTCAACC	600
Db	29756	TTAAAGGAAAGGGAGTTAAAGCGTTTAACTCCCTTTAAACGTGTACAGTCTTTCCACT	29697
QY	601	GCTCTCTTGTAGCTGCTAGCCAGAAACTAGCTGAACTCTAGCTCAGTAAAGTTTTC	660
Db	29696	GCTTTGTAGAACAGCTAGCTAGCGCTTGTGCTTAAGAGTTGGAAGATACACTTCACT	29637
QY	661	GATTCTTACTTGTCCCTAGTGGCAATACAGAGCTGCTGTGATGCACAAAAGAGGACAT	720
Db	29636	GACTTCAAGATTCTTGTGTTGTTAATACCGAAGCTAATATTATGMAAAAGACCGTATC	29577
QY	721	GCTCAGCTCTTGACCGCTCAGCTCAGGAACCGTCTCTTCTATGAAGTATTGGGTC	780
Db	29576	CCAGAACTATTAGCCGCTCAAGTATGAGCGCTGTCTGTTTATGACAGTGTTCGACT	29517
QY	781	ATGCAAGAGCAGGCATAGCAACTTTATCGAGATTGGACCGGGGAAAGTCTTCTCAGGT	840
Db	29516	TTAGTAAAGTGGCATAACAACTTCAATGAGGTAGGACCGAGTTTACAGGT	29457
QY	841	TTTGTAAAAAATTGATCAACTGCTCATTAGCTCATGTGGAAGATCAAGCAGTTTA	900
Db	29456	TTTGTGAAGAAATTGATAAAATTTACTATGTACTAGTGTGGAACATGTTAAGCTTA	29397
QY	901	GTAGCAGCTTTTAAAAAATAG	921
Db	29396	CGTTTATTTTAGATAGAGAG	29376
RESULT 11			
AE010088/c			
LOCUS			
DEFINITION			
Streptococcus pyogenes strain MGAS8232, section 136 of the complete genome.			
1456 bp DNA linear BCT 03-APR-2002			

ACCESSION	AE010088	AE009949
VERSION	AE010088.1	GI:19748917
KEYWORDS	Streptococcus pyogenes MGAS8232.	
SOURCE	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.	
ORGANISM	1 (bases 1 to 14456)	
REFERENCE	Smoot, J.C., Barbican, K.D., Van Gompel, J.J., Smoot, L.M., Chaussee, M.S., Sylva, G.L., Sturdevant, D.E., Ricklefs, S.M., Porcella, S.P., Parkins, L.D., Beres, S.B., Campbell, D.S., Smith, T.M., Zhang, Q., Kapur, V., Daly, J.A., Veasy, L.G. and Musser, J.M. Genome sequence and comparative microarray analysis of serotype M18 group A Streptococcus strains associated with acute rheumatic fever outbreaks	
AUTHORS	Proc. Natl. Acad. Sci. U.S.A. 99 (7), 4668-4673 (2002)	
TITLE	2 (bases 1 to 14456)	
JOURNAL	Smoot, J.C., Barbican, K.D., Van Gompel, J.J., Smoot, L.M., Chaussee, M.S., Sylva, G.L., Sturdevant, D.E., Ricklefs, S.M., Porcella, S.P., Parkins, L.D., Beres, S.B., Campbell, D.S., Smith, T.M., Zhang, Q., Kapur, V., Daly, J.A., Veasy, L. George. and Musser, J.M. Direct Submission	
JOURNAL	Submitted (31-JAN-2002) Laboratory of Human Bacterial Pathogenesis/Rocky Mountain Laboratories/NIH/NIH, 903 S. 4th St., Hamilton, MT 59840, USA	
FEATURES	Location/Qualifiers	
source	1..14456	
	/organism="Streptococcus pyogenes MGAS8232"	
	/strain="MGAS8232"	
	/db_xref="taxon:186103"	
misc_feature	<1..101	
	/note="phi_SDA"	
misc_feature	86..101	
	/note="phi_Sda attC"	
gene	105..917	
	/gene="spYM18_1809"	
CDS	105..917	
	/gene="spYM18_1809"	
	/note="best blastp match gb AAK34481.1  (AE006602) conserved hypothetical protein (Streptococcus pyogenes M1 GAS)"	
	/codon_start=1	
	/transl_table=11	
	/product="conserved hypothetical protein"	
	/protein_id="AAL98329.1"	
	/db_xref="GI:19748918"	
	/translation="MTKKLIAIDLDGLLHNDTISTYTKTIKAVQDKHHVISTGRVRLADYVLOLNKTLPIITFNALDTHPEOKWAYEHNVTLDKYLLRLKLYQDDFQMPFIASEYRKVNVITMTNPESIDPOLFGVDEITDQMALEITKITNPALLMOTHHEDKYLAKNWAPEFKDEIDSMGGPLNILEISSKNVKNKAVNLVLLGIYNMCKDLIAFGDEHNDTEMLAFAGTGAYAMKNSPVLPPYADQQLNFSNEEDGVAKLELFL"	
	1269..2261	
gene	/gene="spYM18_1810"	
	1269..2261	
CDS	/gene="spYM18_1810"	
	/note="best blastp match gb AAK34482.1  (AE006602) mannose-specific phosphotransferase system component IIAB (Streptococcus pyogenes M1 GAS)"	
	/codon_start=1	
	/transl_table=11	
	/product="mannose-specific phosphotransferase system component IIAB"	
	/protein_id="AAL98330.1"	
	/db_xref="GI:19748919"	
	/translation="MGIGIIASHGKPAEGIHQSGLMIFGEQEKVQVTFMNEGPDDLYGHFNNAIQDDADEILVLADWSGPFNQASRVAGENPRKVAITGLNPLMIQAYTERLMDAGAGVEQVAANIKEKSDGIKALPEDLNPIVEATAATEKVNALQGAIPAGTVIGGKGLINLARDVRLHQVATATPKASKADRIIVASDEVAQDRLKQLIKQAPGCVKANVPIISKLEASKDPRFGNTHALILPOTPDALRAVEGGVEINELNVSVAHSGKTMVNNVLSMDKEDVATPEKRLDLGVTFTVRKVPNDKKNLFEIQLKANIK"	
	2350..3159	
	2350..3159	
gene		

```

/ gene="spyM18_1811"
2350..3159
/ gene="spyM18_1811"
/ note="best blastp match gb|AAK34483.1| (AE006602)
putative mannose-specific phosphotransferase system
component IID [Streptococcus pyogenes M1 GAS]"
/ codon_start=1
/ transi_table=11
/ product="putative mannose-specific phosphotransferase
system component IID"
/ protein_id="AAL98331.1"
/ db_xref="GI:19748920"
/ translation="VSDISIIISAILVVIIVIIAFFAGLEGIIDFQFQHOPIVACTLIGLVT
GHLEAGVILGTLQMLALGWANIGAAVAPDAALASVAAAIIMVKSGDFQCKGTRFVYS
TALLVAGLFLMTLVFTLTALVHAGDKAAGNFGAGIERPHEFIALLIGQRIAPVA
TAIVAVTSVQSVLNAMPKLNIEGHCQIGGAVVGVYGVYVNMATREVPWFALGF
ALAAISLTLIAGVIGVIAFIYLNLSKGGNGNAAGADPIGDILEDY"
3176..4087
/ gene="spyM18_1812"
3176..4087
/ gene="spyM18_1812"
/ note="best blastp match gb|AAK34484.1| (AE006602)
putative mannose-specific phosphotransferase system
component IID [Streptococcus pyogenes M1 GAS]"
/ codon_start=1
/ transi_table=11
/ product="putative mannose-specific phosphotransferase
system component IID"
/ protein_id="AAL98332.1"
/ db_xref="GI:19748921"
/ translation="MTSQIKLTSDRQVWRVRSQFIQGSNNYERMQNMGAVALIPAL
KMYTSPEDRAALERMEFNTHPVYAAPIIIGVTLALBEERANGFPIDKAIQGVKI
GMLPLAGIDGPFVFTVTRPILGALGASLASTGNIVGPLLFFFGWNLIRMAFLMYTQE
FYVKGSEITKMSGKGLQDITKGAISLGMFILAIVQVRWVSINFTDILPGKLSQCA
VYVPPDCAVGAEILKTLANAIGMSLIDKVAQTLQGLDLSLIPGLAGLLIFLQNLWL
LKXKVSPIALLIGLFAFGILAHAGIM"
4201..4560
/ gene="spyM18_1813"
4201..4560
/ gene="spyM18_1813"
/ note="best blastp match gb|AAK34485.1| (AE006602)
conserved hypothetical protein [Streptococcus pyogenes M1
GAS]"
/ codon_start=1
/ transi_table=11
/ product="conserved hypothetical protein"
/ protein_id="AAL98333.1"
/ db_xref="GI:19748922"
/ translation="MAQSLNTSVYEYKTKAVSYLGMGKGVGHILLGPKALEFYNDKNVN
DYIQLPHTAINHGANSVRKKVSRHFEIFTDQKFLFASGDSKILKITRQHIGNEKV
ITLPTLMQITFNKFFRH"
4952..6229
/ gene="serS"
4952..6229
/ note="best blastp match gb|AAK34486.1| (AE006602)
putative serYl-tRNA synthetase [Streptococcus pyogenes M1
GAS]"
/ codon_start=1
/ transi_table=11
/ product="putative serYl-tRNA synthetase"
/ protein_id="AAL98334.1"
/ db_xref="GI:19748923"
/ translation="MLDKRKIRDTFTVAAKLKNRVSDTLTHLKELEKERTLLVQ
SEELKARNASAKIAQAQRKEDATQQIADMQKVSADIKTIDNQLVAIDQVADIIIT
VLNPTNDSVPFGADEDNVEIRWGTFRDFFVKAHMDLGEDLDLDMWKGKVTG
ARFLFYKXNLGARDALYNPMLDEHIEKGVQEIITPYMNVHDSMFGTGVQVPKFEITF
ELADNPLVLPFAEVPNTYVGEILLDKELPIYFTAMSPSPRSBAGSAGRTDRLIR
LHOFHKVEMVKFAPEESYOELEKMTANAENILQKLGPRVLSLCTGMGWSAKTY
DLWNIIPAQNTYREIISCSNTEDEFQARRAQIRVRDEADGKVLHLLHTLNGSLAVGRV
AAILNQNEEDGSGVTFEVRKPFMGSEIVISPK"
complement (6462..7232)

```

---

```

/ gene="acca"
/ note="spyM18_1815"
complement (6462..7232)
/ gene="acca"
/ note="best blastp match gb|AAK34487.1| (AE006603)
putative acetyl-CoA carboxylase alpha subunit
[Streptococcus pyogenes M1 GAS]"
/ codon_start=1
/ transi_table=11
/ product="putative acetyl-CoA carboxylase alpha subunit"
/ protein_id="AAL98335.1"
/ db_xref="GI:19748924"
/ translation="WTDVSRVLKEARDQGRITLTDYANLIFFDFMELHGRHFSDDGA
IVGLAVLAGAPVTVIGIKCKNLOHNLARNFGQPNPEGVKALRLKQAEKFRPVV
TFINTAGAPGVGAERGGCGGAIACNLMNSDLKVPITAIIGGGSGGALALAVDQ
VWMLNTWYAVLSPEGFASILWKDGRATEAALMKKITAGELYQMGIVDRIIPBHGIF
SSEIVDIITKANLIBQITSLQAKPLDQLDLDERYQRFKY"
complement (7229..8095)
/ gene="accD"
/ note="spyM18_1816"
complement (7229..8095)
/ gene="accD"
/ note="best blastp match gb|AAK34488.1| (AE006603)
putative acetyl-CoA carboxylase beta subunit
[Streptococcus pyogenes M1 GAS]"
/ codon_start=1
/ transi_table=11
/ product="putative acetyl-CoA carboxylase beta subunit"
/ protein_id="AAL98336.1"
/ db_xref="GI:19748925"
/ translation="MALFRKDKXIRITPNNFLKGSVSHNVPEVPDELFAKCPACKHM
IYKDLGLAKICTCSYNFRISAQERLTLTVDEGSFQELFTSIETKPIRFPQYQEKI
OKKQETGLHVAULTKAMVQGOIATAIMDSHFMASMGTVGKELTRLELAIEEN
LPVIFTAGSARGMQEGIMSLQMAKVSAAVKRHSNAGLFYLTITDPTTGGVTASPA
MEGDIILAEPSQLVGFAGRSRVETVRENLPDDFKAEFLQDHGFVDAIVKXTELDRK
IAHLVAFRGGQ"
complement (8104..9468)
/ gene="accC"
/ note="spyM18_1817"
complement (8104..9468)
/ gene="accC"
/ note="best blastp match gb|AAK34489.1| (AE006603)
putative acetyl-CoA carboxylase biotin carboxylase subunit
[Streptococcus pyogenes M1 GAS]"
/ codon_start=1
/ transi_table=11
/ product="putative acetyl-CoA carboxylase biotin
carboxylase subunit"
/ protein_id="AAL98337.1"
/ db_xref="GI:19748926"
/ translation="MFKKILIANRGEIAVRIIRAARELGISTVAVYSEADKEALHTIL

```

---

```

Query Match      47.5%; Score 437.8; DB 1; Length 14456;
Best Local Similarity 67.2%; Pred. No. 4.9e-109;
Matches 619; Conservative 0; Mismatches 302; Indels 0; Gaps 0;

QY 1 ATGACTAAACAGCCCTTTTATTTGCTGTCGTCAGTGCCCGCAGTATCTAGGATGGACGG 60
DB 13330 ATGACAAAACAGCTTTTATTTGCGCGTCAAGTGCCCAAAATTTAGGATGCAAGG 13271
QY 61 GATTTCTATGATCACTATCCGATTTGTCAGAAAGAGATTGATGAGCGAGTCAGTGCTC 120
DB 13270 GATTTTATGATTACTTTTGCTATTGTAGAAAACCTTTTGATCAAGCTAGTCAAGTATTG 13211
QY 121 GGTATGATTTACGTTATCTCATCGATACGGAAGACAACTCAATCAGACCCGCTAT 180
DB 13210 GGATACGATTTTGGCGCGTTTGATTGATAGTACGAGTTAAACCTTAATCAAACTAGTAT 13151
QY 181 ACGCAACAGCCATCTTAGCAGCTTCGGTTCCTATCTACCGTTTATTTGCAAGAAAGGCG 240
DB 13150 ACGCAACAGCTATTTTGACATCATCAATTCCTATTTACCGGTTTGGGGCTTACATCAC 13091
QY 241 TATCAGCCTGATATGTTGCTGTTTGTCTCTTGAGAAATACTCTGCTTGGTGGCAAGC 300

```



```

/translation="MTRNVISGYGVSAIGNSPEEFNNLKAGKTGIGPIHFDARA
TGISVAEVKPEFDKQKDDARMDTFSIYAVYALDAMWMSGITEENTEDRLGC
IMAGSIGLESQKNSQAIYAKPRKGVAVLYVPLAIANNATGNVALRTGARGVSRAE
VTCAGRANSIGSAPRETKGTADAMTAGGAEEAICELGAGFANULTKREYDPEKA
CIPDKNRGSGVMEGGVLLLESLEHAQAGANILAEIVGYGNTDAPFMTTPSGVG
AENAIKLALEAGASPADVDYVNAHGTSTHANEETESKAIHNVIQKAYVSTKALHG
HALGATAEAVATVQALLNQYAPVNAAGTTELDEGTMINVVLGEGRKADINLALSQDF
GFGGHNVAFAFKWTGE"
gene
3635..4102
/gene="accB"
/ncore="L0187"
3635..4102
/gene="accB"
/ncore="L0188"
/translation="EVIDENCE BY HOMOLOGY BIO07.01 FATTY ACID AND
PHOSPHOLIPID METABOLISM. General. FIRST STEP IN LONG-CHAIN
FATTY ACID SYNTHESIS"
CDS
/transl_table=11
/product="biotin carboxyl carrier protein of acetyl-CoA
carboxylase (EC 6.4.1.2)"
/protein_id="AAK04874.1"
/db_xref="GI:12723695"
/translation="MNISEVKDLMNQFNGSTLREFSKKNADGELSTSKNGOALVGAQ
SAAPTAPFAQAQATPVAESENTEAPVASEAAVEAGEAVTSPLVGAVLKPSPKAEFA
KYGSVVKGTQTLLEIKMKVNEIPAPKDGVIHNVSGEDVIEFGODLMRIK"
4229..4663
/gene="fab22"
/ncore="L0188"
4229..4663
/gene="fab22"
/ncore="L0189"
/translation="EVIDENCE BY HOMOLOGY BIO07.01 FATTY ACID AND
PHOSPHOLIPID METABOLISM. General. FATTY ACID BIOSYNTHESIS"
CDS
/transl_table=11
/product="3-R-hydroxymyristoyl-(3R)-HYDROXYMYRISTOYL-acyl
carrier protein dehydratase (EC 4.2.1.1)"
/protein_id="AAK04875.1"
/db_xref="GI:12723696"
/translation="MTEVNINVTIEALPHRYPPFLVDRVIDIAEDBITAKNVTIN
EEFFQGHFPQIPVMPGVLLIMEAQAAGVLELSKPNKGLVYAGMDNVKRYKKQVTP
GDQVLHAKFKIRGPIAVVEAEATVDGKLAAXGTLPALGR"
5057..6424
/gene="accC"
/ncore="L0189"
5057..6424
/gene="accC"
/ncore="L0190"
/translation="EVIDENCE BY HOMOLOGY BIO07.01 FATTY ACID AND
PHOSPHOLIPID METABOLISM. General. FIRST STEP IN LONG-CHAIN
FATTY ACID SYNTHESIS"
CDS
/transl_table=11
/product="biotin carboxylase (EC 6.3.4.14)"
/protein_id="AAK04876.1"
/db_xref="GI:12723697"
/translation="MFNKKILIANRGEIAVRIIRAARELGVGTVAIXYSTDARDLHVHL
ADEALCVGPASTESYNNMNLISAAVTAQAIHPGFGFLSENSFARCEEMNIFK
IPSAKMDMMGNVREILVPAFEAQAQAQAFNGAMFIERMIFPARIHLEVLIGDSQ
AGGGKGRKVVNVEELVPAFEAQAQAQAFNGAMFIERMIFPARIHLEVLIGDSQ
GNVHLGERCSLQNNQKVLSESPVAIGHTLRNEITAAVAAQAHVSEVNAETIEP
LDESCKGFMEFNMVQVHEVPVFVSGDIVKEQIRVANGNELSVKQEDVTFSGH
AIECRINAENPKFNFPSPGKTNLFLPSGGVGLRVDMSAYSGISYIPPYDMSIAKVI
VHGNRFEALMKMORALEFDEGVITNVDFOLELISDPHVVAAGDYDTSPLGNVFLPE
YQKED"
6510..7376
/gene="accD"
/ncore="L0180"
6510..7376
/gene="accD"
/ncore="L0181"
/translation="EVIDENCE BY HOMOLOGY BIO07.01 FATTY ACID AND
PHOSPHOLIPID METABOLISM. General. FATTY ACID BIOSYNTHESIS"
CDS
/transl_table=11
/product="acetyl-CoA carboxylase carboxyl transferase

```

```

subunit beta (EC 6.4.1.2)"
/protein_id="AAK04877.1"
/db_xref="GI:12723698"
/translation="NALFKKKYIKINPNRSIIIEKQARQEPVDPDLFAKCPACKHTIY
QKDLGKVCNCDYNFRITAKELAIKADKDSFVEMFTGIESKNPLDFPGYPEKLA
TKARGLDEAVITGATIKGQKTALAIINDSTFINASMGTVVGEKILTRIFEVATPKLP
IIVFTASGARMQEGIMSLMQMAKTSAAVKRHSNAGLFYITVLTPDTTGGVTASPASL
GDILILAPQSLLIGFAGRRVIEQTVRQTLDPDFQKAEFLNHGFVDIAIKVRELQKLA
LLELHTEVENV"
gene
7369..8151
/gene="accA"
/ncore="L0190"
7369..8151
/gene="accA"
/ncore="L0191"
/translation="EVIDENCE BY HOMOLOGY BIO07.01 FATTY ACID AND
PHOSPHOLIPID METABOLISM. General. FIRST STEP IN LONG-CHAIN
FATTY ACID SYNTHESIS"
CDS
/codon_start=1
/transl_table=11
/product="acetyl-CoA carboxylase carboxyl transferase
subunit alpha (EC 6.4.1.2)"
/protein_id="AAK04878.1"
/db_xref="GI:12723699"
/translation="MSEIIDSILTGEVSQKNIADIINEARAADRISARDIIDKVFDTDF
ELHQRQITDDDEALVGGIARFNQAVTVIGQKGNLAENLATNFGQSPNGRYKALR
LMQKREKFRPVITFINAGYAPGIEAEEGCGEAIARNLLEMSDLKPIIAFIATFEG
GSGGALALANKVYILENAVSVLSPEGFATILWGDGSRDRDEAELMKITAPHLLEM
GLVDIAIIEENLKRQITSLGDSLLTETELAEQRYERFRKY"
8280..9422
/gene="metB2"
/ncore="L0191"
Query Match 42.2%; Score 388.8; DB 1; Length 10701;
Best Local Similarity 64.6%; Pred. No. 1.4e-95;
Matches 598; Conservative 0; Mismatches 322; Indels 6; Gaps 1;
QY 1 ATGACTAAACAGCGCTTTTATTCGTGTCGTCAGGTGCCAGTATCTAGGATGGACGG 60
DB 615 ATGACTAAACAGCAATTTTATTCAGTCAAGGGCACAAAGCTTGGATGGACGT 674
QY 61 GATTTCTATGATCAGTATCCGATTTCTCAAAGAAACGATTTGATCGAGCGAGTCAGTGTC 120
DB 675 GACTTATATGACCAATATGAAACAGTTAAAGCAACTTTTGTATGAAGCAAGTCAAGCTTTA 734
QY 121 GCTTATGATTTACGTTATCTCATCGATACGGAGAGACAACTCAATCAGACCCGCTAT 180
DB 735 GGAATGATTTTCGAGCTTTGATTTGATGATGAAGAAACCTAAATGAACAAAGTAC 794
QY 181 ACGCAACAGCGCATTTCTAGCGACTTCGGTTGCTATCTACCGTTTATTTCAGAAAGGCG 240
DB 795 ACTCAACCTGCAATTTTAAACACCTCTGTGCTATTTTACGTTTGTAAAGTGAATGG 854
QY 241 TATCAGCTGATATGTTGCTGTTTGTCTCTTGAGAAATACTCTGCGCTTGGTGGCAAGC 300
DB 855 ATTAAACCTGACCTTGTGCTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTATCA 914
QY 301 GCGCGCTTGGATTTTGAAGATGCGGTTGCTTGGTAGCTAAAGCGTGGAGCCTATATGAA 360
DB 915 GGAATCATTTGATTTTCAAGAGCAGTTAAAGCTTGTAGCTAACGTTGGTCAATATGACA 974
QY 361 GAAGCGGCTCTGCTGATCTCTGGCAAGATGTTAGCAGTTCTCAATACCGCAGTAGAGTC 420
DB 975 GAAGCTGACCGACTGCTTCTGTTAAATGGTTCGCTGCTGCTGCTGCTGCTGCTGCTG 1034
QY 421 ATTGAAGAGCGCTGCAAAAGCGCTTCTGAACCTT-----GGAGTGGTTACTCCAGCCCAAC 474
DB 1035 ATTGAAGAAATTTGCCAAGAGCGCCGCAATTTAAAGCGCGTATTGTGAGTCCGCAAT 1094
QY 475 TATAACACACCTTGCAAAATCGTCAATTCGTGAGAAAGTGGTTGCTGAGTGTGAGCGGTT 534
DB 1095 TATAACACCGCCGACAAATTTGTTTGTGTTGAGTTGAGCGGTTGATTATCTGCTGT 1154
QY 535 GAACCTTTTGCAGAGAGCAGGTGTCACAAACGCTTGATTCCTCTTAAGGTCTCAGTCCCTTT 594

```

```
Db 1155 GAGTGTCTAAAGAGCCGAGCTTCGTAACTATTGAATTAATAAGTTTCAGGACCTTTC 1214
QY 595 CACACCGCTCTCTTGGAGCTCTAGCAGAACTAGCTGAACTCTAGCTCAGTCAAGT 654
Db 1215 CATACAGCAATTTTAAACACGACACTGNAATTTGGCTTGGAGCTTGATAAATGAT 1274
QY 655 TTTTCAGATTTACTTGTCCCTAGTCGGCAATACAGAGCTGCTGTGATGCAAAAGAG 714
Db 1275 TTTAGACCTTTTGAATTAGCATTATCTCAATACGAGTCTGCTGAAAGTAAAGAAATGAT 1334
QY 715 GACATTGCTCAGCTCTTACGCGCTCAGGTCAGGACCGCTTCGTTCTATGCAAGATT 774
Db 1335 GAAGTCAAGGACTTTTACGCGCTCAGTTCATGGAACAGTTCTGTTTATGATCGGTT 1394
QY 775 GGGGTATCAGAGACGAGGCATAGCAACTTTATCGAGATGGACCGGGAAGTCTTG 834
Db 1395 GAAACAATGCAAAACTAGGGCGACTCGCTTTATGAATGGTCTCGGAGAGTACTT 1454
QY 835 TCAGGTTTGTGTTAAATAATGATCAAACTGCTCACTTAGCTCATGTCGAGATCAGCG 894
Db 1455 TCAGGTTTCATTAAATAATGATCAAACTGCTCACTTAGCTCATGTCGAGATCAGCG 1514
QY 895 AGTTTAGTAGCAGCTTTTGAATAATA 920
Db 1515 TCATTGGAAGCTTGATTAAATCAGTA 1540

RESULT 13
AE004276/c
LOCUS
DEFINITION
14530 bp DNA linear BCT 31-JUL-2000
Vibrio cholerae chromosome I, section 184 of 251 of the complete
chromosome.
ACCESSION AE004276
VERSION AE004276.1
KEYWORDS GI:9656555
SOURCE
ORGANISM
Vibrio cholerae
Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE
1 (bases 1 to 14530)
AUTHORS
Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L.,
Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Unayam, L.,
Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,
Ermolaeva, M.D., Vamathevan, J., Baas, S., Qin, H., Dragoi, I.,
Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D.,
Nierman, W.C. and White, O.
TITLE
DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae
JOURNAL
Nature 406 (6795), 477-483 (2000)
MEDLINE
20406833
PUBMED
10952301
REFERENCE
2 (bases 1 to 14530)
AUTHORS
Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L.,
Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Unayam, L.A.,
Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,
Ermolaeva, M.D., Vamathevan, J., Baas, S., Qin, H., Dragoi, I.,
Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D.,
Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R.,
Mekalanos, J.J., Venter, J.C. and Fraser, C.M.
TITLE
Direct Submission
JOURNAL
Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1..14530
/organism="Vibrio cholerae"
/strain="N16961"
/serotype="O1"
/db_xref="taxon:666"
/chromosome="I"
/notes="biotype: El Tor"
223..1806
/gene="VC2012"
223..1806
CDS
gene
223..1806
CDS
223..1806

/genes="VC2012"
/notes="similar to GB:AE000782; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="sodium-dependent transporter"
/protein_id="AAF95160.1"
/db_xref="GI:9656556"
/translation="MDDRKVRANIVRTLPFPDPSATADEKLTHTLGYKELPMKR
EWGSRVGVLAAGSAIGLNIWPPFYAYGGGAFPIYLFAMVTAGIYFPMILEF
SGDKHRSAGPTLAKHSKPEWLGWFOVIGIAAVIAYVAVIGWISYFGMSFTQSW
GSDTNAVFTSEVYKLGDSNPTNLGSIOMHIAFLIAVITVAALVGVKKGIERASK
IAMPILFINVILLIARMIFLPCALDGVNMPQDFSKTDVWVAAVQOIFRTEISIG
FAMLAISYLEKSDITNNATFVILGPFVLAGIMIFSVLWAGQGLPEHVV
SAGVGLAFTLPAAILNDIPYILGFLPLALVAGLSSHISIMEAVTSIDLKLGWS
RKAANIVIGFLVSMFATNGLLLDLVDFANNVIMASCLVELVMTWLKIS
DYKAVNS"SDPSVGVWFDICLRFVSPVILAIIVATKIQALFTGYGGYDLTGLWAIL
AALLVIGILINASSRKEY"
complement (2019..3521)
/gene="VC2013"
complement (2019..3521)
/gene="VC2013"
complement (2019..3521)
/notes="similar to GB:U02618 SP:P05053 PID:147393 GB:U00096
PID:1651541; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="PTS system, glucose-specific IIBC component"
/protein_id="AAF95161.1"
/db_xref="GI:9656557"
/translation="MLTRELGYVSPNSNNLIRISINFMKAFANLQVKGKALMLPV
SVLPVAGILLVGAANFSLPEVSHLMQAGSVFGQMLLFAVGLVFNNDGVS
GSAIVGVYIVATLKVATVGVSGIDTGLVGLAGVAAANFNRFKIQLPYLG
FPAGKEAVIITGTISIALGVSLFIPWPGIAIATFSDAANQDPVAFGYGVIVER
SLTPEGLHHIWNVPFYVQAGTCVNGAGTVMCTELTADDAAGNGFQGLAGGY
LPMFGLPAAFAIAHCAPENRKYMGIMASALTSFLTGITETIEPAFLVAPVLY
AIHVLGLAYVLTNALGVVHGTFSNGFIDFVQSPRADNMLLVGLSGIGFVAVLYV
VFTFVIRALNLTGKPGEDSADKSGNELAGDLVAAFPGGKANINDADCTRLEVS
VADTALVDQDKLKLGAGVVMVSGVQAIFGKSDNLKTNDEWIRNNG"
complement (3893..4660)
/gene="VC2014"
complement (3893..4660)
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF95162.1"
/db_xref="GI:9656558"
/translation="MFVDSHCHLDKLDYDLHADYSDVLDKAKHAKVEYLLSGVTLT
SFKMLAMLEPFSQVASCVHPLDVQSDLSLPHQAQHRVVAIGTGDIYHQY
ETADLQKLRFOQIETAVALKPLIHTNNACDQLDLREGKACQCGSVIHCFTEDL
AFAQAMELGFIYSIGVITFQATDLKEVVKALPLDLRLIETDSPYLAPVPHRGKEN
QPAYVVEVAAVIAALKGVSLTEVAQKTTQNQNFLR"
complement (4651..5625)
/gene="VC2015"
complement (4651..5625)
/notes="similar to GB:L01483 SP:P28631 GB:L04577 PID:145783
PID:145799; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="DNA polymerase III, delta prime subunit"
/protein_id="AAF95163.1"
/db_xref="GI:9656559"
/translation="MSNUNSLYVLPVWQVQAGLAAGKISSATLIQASEGVGVESL
VELMARTLMCTSSQSEPCGCHSCGLMGSNHPFHVVKPEKIGKISITVEQIRQWRI
AESQSLGRLIVIEPAENSSANLKTLEPAPNCLFLVTSIRIKHLLTIVS
RCORLVLPATPALVVEMLKGGCITPAYALHLCADSPKTRAFLEGAEGKHYELES
QGMNALSQDVNAQLKCIALIDADLTTHLVWVCLTDQAKHFGVQOQVYPASALA
GRFTYSLKHVQTSLSERLEMEQLQFSGLENTLELLQLWLYQFSDDEETCL"
complement (5618..6256)
/gene="VC2016"
complement (5618..6256)
```

```
/gene="VC2016"
/notes="similar to GB:L42023 SP:P44719 PID:1003796
PID:1222388 FID:1204706; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="thymidylate kinase"
/protein_id="AAF95164.1"
/db_xref="GI:9656560"
/translation="MRAKIVIEGLEAGKSTAIQVVTETLQONGIDHITRERGGT
LLAKRUALVKEHPGEEELQDITTELLVTAARVQLVENVIKPALARGVWVGDRHDS
SQYGGGRIAPSTWQSRLDGLFDPDLTLIDIDPKLGLERARGELDRTERM
DISFFRAREYELANSDDSVVIDAQSISQVTADIRRALQDWLSQVNRV"
/complement(6253..7269)
/gene="VC2017"
/complement(6253..7269)
/gene="VC2017"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF95165.1"
/db_xref="GI:9656561"
/translation="MVKLVLVVALIGIVAGSYFYVVKMDQYLAOPLMIOEAGLVT
IASGTLRELQGLTEQAMQDSFVAEWRRHPELSKTAKGTIKLOPDMSEUQALAL
LVSGKHQSITVEGSRFOEWRDILASNETIQTLTGLTEIDIAKALGIEHEKLEGL
FTAETHFTKGTSDVRLKRAKQLEKFLQVWEHQEGLPIQTPEALILASIIEXE
LSAEBERETAAVFIRNLKRMELQDPTVIYMGSETYDGNIRKRLARTPYNYVI
NGLPPTPLAMPGEASIALNPQSDLYLFVVASGEGGHFNFSKTLADHNRVAYLKKL
RTKQ"
/complement(7266..8099)
/gene="VC2018"
/complement(7266..8099)
/gene="VC2018"
/notes="similar to PID:1173841 PID:1173844; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="4-amino-4-deoxychorismate lyase"
/protein_id="AAF95166.1"
/db_xref="GI:9656562"
/translation="MSRFTLWSESMYVNGQRNRPVPIHRSFOYGDGFTILTKEG
QVQWQSHKRAQACLDLIHPNPWDRVQGLQSMILPOEKAGLKHISRLGGRGY
SPTQVESIVTISAPFAHPYQAWRDKGLAVGICQORMGLNPLLAGHKHNNLEQILL
KXEMDNAGDDGVCIDINKVETTAANTFWCRDGTMTPTCLRHAGVAVARRQLEL
AQQEPIVIDEFTLEDLSAREEVFIINALLEVAVPVIGQQORLTIGSMTRRFQESSN
S"
/complement(8147..9391)
/gene="VC2019"
/complement(8147..9391)
/gene="VC2019"
/notes="similar to PID:1173841 SP:P55338 PID:1173843;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="3-oxoacyl-(acyl-carrier-protein) synthase II"
/protein_id="AAF95167.1"
/db_xref="GI:9656563"
/translation="MSKRVRVVTGMCLSPVCNTVSESSWKALLAQSGIVNIEHPDIT
NSTRPAGLVKFDCEQKNSKDKARDKDLFIQIGIAAGIQALEDSGLEVNEENRARG
VAIGSGLELIETQHALIEKPRKVPFFVPTSTVNMIGNLSIMRGLRGNIAI
STACTGLHNIGHARMIAIYGDADAVAGGAEKASTPLMGAGFAAKALSTRNDEPQK
ASRPMDKRDGTVLGDAGIIVLEVEYERKARGAKIYAEVVGFMGSDAYHMTSPED
SGGALAMAAARDAGTCEQIGYVNAHGTSTPAGDVAEVKGIKRALGEAGTKOVLVS
STKMTGTHLLGAAGSVEAIIITWMSLVQDVQVPTINLDNPEELGVDLVPHVARKVESM
EYAMNCSFGGCTNGSLIFKRM"
/complement(9480..9812)
/gene="VC2020"
/complement(9480..9812)
/gene="VC2020"
/notes="similar to PID:1173841 SP:P55337 PID:1173842;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="acyl carrier protein"
/protein_id="AAF95168.1"
/db_xref="GI:9656564"
/translation="MVRPETSRQLSLVRINYGIALGEICKGKMSNIEBVRVKLI
VEQLGVDEABEVKNSESPVEDLGADSLDTVELWALBEEFOTEIPDZAEKITTQAAI
DYVTSNAQ"
/complement(9952..10698)
/gene="VC2021"
/complement(9952..10698)
/gene="VC2021"
Query Match 23.8%; Score 219.6; DB 1; Length 14530;
Best Local Similarity 54.5%; Pred. No. 3.9e-49;
Matches 507; Conservative 0; Mismatches 414; Indels 9; Gaps 3;
QY 1 ATGACTAAACAGCGCTTTTATTTGCTGGTCAAGTCCCAAGTATCTAGGATCGGACGG 60
DB 11637 ATGAGTAAGTTTGTATCGTATTTCCAGGTCAAGGCTCGAAGCAGTAGGTATCTGCTGCT 11578
QY 61 GATTTCATGATCAGTATCCGATTGTCAAGAAACGATTGATCGAGCGAGTCAAGTGTGCTC 120
DB 11577 GACCTTGGCGAGAGTATGCTGTGTTAAACAACATTCGCCGAAGCTTCAGAAGTGTCT 11518
QY 121 GGTATGATTTACGTTATCTCATC---GATACGGAAGAACACAACTCAATCAGACCGCG 177
DB 11517 GGTTACGATCTGTGGCGCTGGTTCAAGATGGCCCTGTGGAAGATCTCAACCAAACTTTC 11458
QY 178 TATACGCAACAGCACTTCTAGCGACTTCGGTTCGTATCTACCGTTTATTCAGAAAG 237
DB 11457 CGTACTCAACTGCTGTGCTTGGCGCTCTGTTCGGATTGGCGGTATGCGCAGCACTG 11398
QY 238 GG---CTATCAGCCTGATATGGTTGCTGGTTGTCTCTTGGAGAAATCTCTGCGCTTGGTG 294
DB 11397 GGTCTTGAGCAACCTCGCGTTTATAGTGGTCACAGCTGGTGGTGAATATTCAGCACTGTA 11338
QY 295 GCAAGCGCGCTTGGATTTTGAAGATGCGGTTCCTTGGTGTAGTAGCGTGGAGCGCTAT 354
DB 11337 TGTGCGGCGGTGATTGATTTTAAACAAGCGATCAAGCTGGTTGAGCTGCGGTGTCATTTG 11278
QY 355 ATGAAAGAAAGCGGCTCTGCTGCTGACTTGGCAAGATGTTAGCAGTCTCAATACGCCAAGTA 414
DB 11277 ATGCAACAAGCGGTGCTCGCGGTACGGTGCATATGACCGGATCATCGGTCTAGAAGT 11218
QY 415 GAGTCTATTGAAGAAGCGCTGTCAAAAAGCTTCTGAACTTGGAGTGTGTACTTCCAGCCAA 474
DB 11217 GAAGCGATTGTAAAGCGCTGTGCTGACCGCGCGCAAGGTGAAGTGGTTTCTCTGTAAAC 11158
QY 475 TATAACACCTGCACAAATCGTCAATTCCTGGAGAACTGTTGAGTTCAGTTCAGGCGGTT 534
DB 11157 TTTAACTCACAGGCCAAGTGGTTATTGCTGGTCAAAAAGATGCGGTTCAGCGTGGCGGC 11098
QY 535 GAATTTTGAAGAAGCAGGTGCGCAACGCTTGATTCCTCTTTAAGTGTTCAGGTCCCTTT 594
DB 11097 GTTCTGTGTAAAGAAGCGGCGGCAACGCTGCGCTCTGCGCAGTTTCCGTACCATCA 11038
QY 595 CACACCGCTCTCCTTGTAGCGCTGTAGCCAGAAATAGCTCAAACTCTAGCTCAGGTAAGT 654
DB 11037 CACTGCGGTTGATGAAGCTGTGCGCGATGAATTTGCAAAAATCTTAGCAGAGCTTGAA 10978
QY 655 TTTTCAGATTTTACTTGTCCCTCTAGTCGGCAATACAGAAGCT---GCTGTGATGCAAAA 711
DB 10977 TTCAATGCACCAAAATTCGGGTTCATCAATAACGTTGATTTGCTGGCTGAAACGGATCCG 10918
QY 712 GAGACACTTCTCAGCTCTTTCACGCGTCAAGTCAAGAACCCCTGCTGTTTCTATGAAGT 771
DB 10917 GTAAAAATTAAGATGCGTTGATTTCGTAACCTTATAGCCAGTTCGTTGGACTGAATGC 10858
QY 772 ATTCGGGCTATGCAAGAACGAGGCATAAGCAACTTTTTCAGAGATTGACCGCGGAAAGTC 831
DB 10857 GTTGAAACAATGAGCGCACAGGTGTGCAAAAGCTGATTGAATGGGCGCGGTAAAGTA 10798
QY 832 TTGTCAGGTTTTGTTAAAAAANTTGATCAAACTGCTCACTAGCTCATGTGGAAGATCAA 891
```

```

Db      10797  TTGACTGGTCTACAAACGTTTGTAAACCCCTAGAGGTGTCGCAGTCAATGACGTG 10738
QY      892   GGAGTTTAGTACACTTTTACAAAAATAG 921
Db      10737  GCTTCTTGGATGCGGTGAATAACAACAG 10708

RESULT 14
LOCUS   AE013105/c
DEFINITION
Thermoanaerobacter tengcongensis strain MB4T, section 132 of 244 of
the complete genome.
ACCESSION
AE013105 AE008691
VERSION
AE013105.1 GI:20516482
SOURCE
Thermoanaerobacter tengcongensis.
ORGANISM
Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
Thermoanaerobacteriaceae; Thermoanaerobacter.
REFERENCE
1 (bases 1 to 11845)
Bao,Q., Tian,Y., Li,W., Xu,Z., Xuan,Z., Hu,S., Dong,W., Yang,J.,
Chen,Y., Xue,Y., Xu,Y., Lai,X., Huang,L., Dong,X., Ma,Y., Ling,L.,
Tan,H., Chen,R., Wang,J., Yu,J. and Yang,H.
A Complete Sequence of the T. tengcongensis Genome
Genome Res. 12 (5), 689-700 (2002)
21992816
11997336
2 (bases 1 to 11845)
Bao,Q., Xu,Z., Hu,S., Dong,W., Chen,Y., Wang,J., Yu,J. and Yang,H.
Direct Submission
Submitted (07-OCT-2001) Beijing Genomics Institute/Genomics and
Bioinformatics Center, Institute of Genetics and Development,
Chinese Academy of Sciences, Beijing Airport Industrial Zone B8,
Beijing 101300, China
3 (bases 1 to 11845)
Li,W., Xuan,Z., Yang,J., Ling,L. and Chen,R.
Direct Submission
Submitted (07-OCT-2001) Institute of Biophysics, Chinese Academy of
Sciences, Beijing 100101, China
4 (bases 1 to 11845)
Tian,Y., Xue,Y., Xu,Y., Lai,X., Huang,L., Dong,X., Ma,Y. and Tan,H.
Direct Submission
Submitted (07-OCT-2001) Institute of Microbiology, Chinese Academy
of Sciences, Beijing 100080, China
location/Qualifiers
1. 11845
/organism="Thermoanaerobacter tengcongensis"
/strain="MB4T"
/db_xref="taxon:119072"
ccomplement(76. 1153)
/gene="ELP3"
/note="TIE1468"
complement(76. 1140)
/gene="ELP3"
/note="Best Blastp hit = gi|7462101|pir||A72295 conserved
hypotheical protein - Thermotoga maritima (strain MSB8)
gi|4981648|gb|AAD36179.1|AE001769.9 (AE001769) conserved
hypotheical protein [Thermotoga maritima], score 202,
E-value 6.00E-51"
/codon_start=1
/product="3-oxoacyl-
/transl_table=11
/note="Pfam match to entry ketoacyl-synt C, Beta-ketoacyl
synthase, C-terminal domain, score 238.8, E-value
7.708-68"
protein_id="AA024690.1"
/gene="FabB"
complement(2347. 3078)
/misc_feature
/note="FamB"
/note="Pfam match to entry ketoacyl-synt, Beta-ketoacyl
synthase, N-terminal domain, score 295.3, E-value
7.708-85"
complement(3088. 3092)
/gene="FamB"
complement(3111. 3127)
terminator

```

```

EVEKEGKGGKISIFCNERKLSITVIGQKYNKRFLDEKYGLQVQVKKLEKIEEKIV
LSYDKLYPISITDFIKONF"
complement(1091. 1108)
/gene="ELP3"
/note="putative rho-independent transcription terminator"
complement(1137. 1837)
/gene="Rnc"
/note="TIE1469"
complement(1137. 1823)
/gene="Rnc"
/note="Best Blastp hit = gi|10175109|dbj|BAB06208.1|
(AFO01515) ribonuclease III [Bacillus halodurans], score
193, E-value 1.00E-48"
/codon_start=1
/transl_table=11
/product="dsRNA-specific ribonuclease"
/protein_id="AA024691.1"
/db_xref="GI:20516484"
/translation="MLTYLEQKINYEFDKDTLLLEALTHSSWAHEGKNEKVNLERLEF
LGDSVLSLVISYLYKNRKDEGLSKYRAIVCEPSLARCARKEILGSLQFMRMGKE
EISGGRDRDSILADAMEALLAAVYLDGGLAEVRRVILDLFKEIIDVLKGIYRDYKT
RIQEVYQSMNEVGKITYELVEBIGDPHNKTFVTQVKIGDVVLGIGQGSKESEQAAM
EALSUKGLIK"
complement(1149. 1153)
/gene="ELP3"
complement(1152. 1352)
/gene="Rnc"
/note="Pfam match to entry dsrm, Double-stranded RNA
binding motif, score 85.2, E-value 1.30E-21"
complement(1434. 1706)
/gene="Rnc"
/note="Pfam match to entry Ribonuclease_3, RNase3 domain.,
score 139, E-value 8.40E-38"
complement(1833. 1837)
/gene="Rnc"
complement(1849. 3092)
/gene="FamB"
/note="TIE1470"
complement(1849. 3078)
/gene="FamB"
/note="Best Blastp hit = gi|7433750|pir||G69842
3-oxoacyl-[acyl-carrier-protein] synthase (EC 2.3.1.41) -
Bacillus subtilis gi|2633471|emb|CAB12975.1| (Z99109)
similar to 3-oxoacyl- acyl-carrier protein synthase
[acallus subtilis] gi|2633488|emb|CAB12991.1| (Z99110)
similar to 3-oxoacyl- acyl-carrier protein synthase
[Bacillus subtilis], score 496, E-value 1.00E-139"
/codon_start=1
/transl_table=11
/product="3-oxoacyl-(acyl-carrier-protein) synthase"
/protein_id="AA024692.1"
/db_xref="GI:20516485"
/translation="MNRVVVTGLGVISPLGNTVERFNSLINGENGIGLVTKFDTSF
PTKVAAEVKDFDFTLYIDKKEARRMDRFIQFAVASTKLAEDAGIELDKVDLNRFGVI
YGTGGTIEFENQKMYKPKSVSPFFIPMIMIANMAAGQIATFGLKGINETIVN
ACGAE3NATGEAFKAIQRGDADLIVTGGTAAITVPFAMFCAMKAMTNDPDPSTASR
PDLANEDCFVMGEGSATILLESLEHAQKRAKIVAEIIGYGATDAVHITAPAGEG
ARAMEALDKGVSYDMVDVINAHGTSTENYDKETMALIKVFKDHAYKLCVSSNKS
MIGHLGAGAVEAVATVDTINHGVTINQPTDPEDCLNYPNKAIERENVVAIS
NSFGFGHNATLLFFK"
complement(1855. 2325)
/gene="FamB"
/note="Pfam match to entry ketoacyl-synt C, Beta-ketoacyl
synthase, C-terminal domain, score 238.8, E-value
7.708-68"
complement(2347. 3078)
/misc_feature
/note="FamB"
/note="Pfam match to entry ketoacyl-synt, Beta-ketoacyl
synthase, N-terminal domain, score 295.3, E-value
7.708-85"
complement(3088. 3092)
/gene="FamB"
complement(3111. 3127)
terminator

```



TITLE Methods for monitoring multiple gene expression

JOURNAL Patent: WO 0229113-A1 1143 11-APR-2002;

Novozymes Biotech, Inc. (US) ; Novozymes A/S (DK)

FEATURES Location/Qualifiers

source

1. .950

/organism="Bacillus licheniformis"

/db\_xref="taxon:1402"

BASE COUNT 276 a -223 c 259 g 192 t

ORIGIN

Query Match

Best Local Similarity 22.1%; Score 204; DB 6; Length 950;

Matches 474; Conservative 0; Mismatches 385; Indels 9; Gaps 3;

QY 1 ATGACTAAACAGCCCTTTTATTGTGTCAGGTGCCAGTATCTAGGGATGGAGCG 60

Db 1 ATGGCAAGATTGCTTTCTATTCCGGGCCAAGGTCGACATATCGCATGGACAC 60

QY 61 GATTTCATGATCAGTATCGGATGTCAAGAAACGATTGTCGAGCGAGTCAGGTGCTC 120

Db 61 GAATTGTATGAAAAGAACCGGATGCGAAGAGATTTTGAAGAAGCGGATCAAAACGCTT 120

QY 121 GGTATGATTACGTTTATCTCA---TCGATACGGAAGACAACTCAATCAGACCCGC 177

Db 121 GAACAAACTGACACCCCTCATGTTGAGGGGATGCAAGAACTGACGTTACATAC 180

QY 178 TATACGCAACACCCATTCTAGGACTTCGGTTGTATCTACCGTTTATTGCAAGAAAG 237

Db 181 AACGGCAGCCAAAGCCCTTTTAACGGCGAGCATCGACGCTTGAAGAACTGAAGGAATAC 240

QY 238 GGTATCAGCCTCATATGTTGCTGTTGTTGTCTCTGGAGATATCTCTGCTTGGTGCA 297

Db 241 GGCATTAAACCCACTATCGGCGAGTACAGCTCGGCGAATACAGCGCATTTGGTCGT 300

QY 298 AGCGGCGCTTGGAATTTGAAGATGCGTTGCTTGTAGCTTAAGCTGAGCCCTATATG 357

Db 301 GCGGCGCTTGTGCTTTAAAGATGCGTTTATGCGTCAGAAAGCGCGGCAATTCATG 360

QY 358 GAAGAGCGGCTCCTGCTGACTCTGCGAGATGGTAGAGTTCTCAATACGCCAGTAGAG 417

Db 361 ATGAAAGCGGTGCGGCGGAGAGCGCGATGGCGCCATTCTCGGATGGACGCCAG 420

QY 418 GTCATTGAAGAGCCCTGTCAAAAGCTTCTGAACCTTGA---GTGGTTACTCCAGCCAAC 474

Db 421 GCGCTGAAGAGAGTGAAGCGACAAAATTTCCGAAGAGGAACCTTTGTCAGCTGCCAAT 480

QY 475 TATAACACAGCTGCACAAATCGTCATTGCTGGAGAGTGGTTGCAAGTTGATCGACCGGTT 534

Db 481 TTGAACCTGCCCTGGGCAAAATCGTCATCTCGGAACAGCTTAAAGGCGTGGAGCTCGCTTCA 540

QY 535 GAACTTTTGCAGAGCAGGTGCGAAAGCTTGTATTCCTTTAAGGTGTCAGGTCCCTTT 594

Db 541 GAGCTTGCAGAGAGAGGCGCAAAACGCGGATTCCTCTCGAAGTACGCGGCGGCTTC 600

QY 595 CACACCGCTCTCTTGAAGCTGCTAGCCAGAACTAGCTGAACCTCTAGCTCAGGTAAGT 654

Db 601 CATCTGAGCTGATGAAGCGGCGAGCTGATGAAGCTTCGTGAAGTCTTTCATGCGTGCAG 660

QY 655 TTTTCAGATTTTACTTGTCCCTAGTGGCAA---TACAGAGCTGCTGTATGCAAAA 711

Db 661 ATCAAGCAGCGAGCCATTCGGTCTGCTCCAAAGTAAAGCGGCGACTTTGTACGGATAAA 720

QY 712 GAGGACATTGCTCAGCTTTGACGGTCAAGGTCAAGAACCCGTTTCGTTTCTATGAAAGT 771

Db 721 GACGACATTGAAGATAAAGTGAACAGCTGTATTCCTGTTACGCTTTGAAGAAACA 780

QY 772 ATTGGGGTCATCAAGAACGAGCATAGCACTTTATCGAGATTGACCGGGAAGTC 831

Db 781 ATCAGCGGCTGATTGACGAAGCGGTACGACCTTCATTGAATCGTCCCGGAAAGGTT 840

QY 832 TTGTCAGGTTTGTAAAAAATTGATC 859

Db 841 TTGTCAGGCTTGTGAAGAAAGTGAACC 868

Search completed: June 11, 2003, 18:44:52  
Job time : 2441 secs

GenCore version 5.1.6  
 Copyright (c) 1993 - 2003 Compugen Ltd.  
 OM nucleic - nucleic search, using sw model  
 Run on: June 11, 2003, 17:03:56 ; Search time 264 Seconds  
 (without alignments)  
 7856.403 Million cell updates/sec

Title: US-09-308-397-1  
 Perfect score: 921  
 Sequence: 1 atgactaaacagcctttt.....tagcacttttagaaaaatag 921

Scoring table: IDENTITY\_NUC  
 Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_101002.\*  
 1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
 2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
 3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
 4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
 5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
 6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
 7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
 8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
 9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
 10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
 11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
 12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
 13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
 14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
 15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
 16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
 17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
 18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
 19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
 20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
 21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
 22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
 23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
 24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	921	100.0	921	19	AAV37239
2	913	99.1	921	22	AAH90730
3	913	99.1	19702	19	AAV52140
4	911.4	99.0	921	23	AAS55845
5	911.4	99.0	3580	19	AAV65243
6	908.4	98.6	945	22	AAH90843
7	529.4	57.5	547	19	AAV37241
8	523.2	56.8	1196	19	AAV38542
9	523.2	56.8	1199	18	AAV30801

C	10	523.2	56.8	1209	19	AAV37406	Streptococcus pneu
	11	439.4	47.7	942	24	ABN68641	Streptococcus poly
	12	402	43.6	636	18	AAV30880	Streptococcus pneu
	13	388.8	42.2	2365589	24	ABA90521	Genomic sequence o
	14	387	42.0	924	24	ABN68640	Streptococcus poly
	15	387	42.0	2155561	24	ABN71527	Streptococcus poly
	16	366	39.7	450	19	AAV37240	N-terminal coding
	17	344.4	37.4	942	23	AAS53186	Enterococcus faeca
	18	340.4	37.0	3656	20	AAV13514	Enterococcus faeca
	19	204	22.1	950	24	ABK73852	Bacillus lichenifo
	20	204	22.1	960	24	ABK73795	Bacillus lichenifo
	21	196.4	21.3	939	23	AAS53241	Haemophilus influe
C	22	196.4	21.3	1830121	17	AAV42063	Haemophilus influe
	23	185.6	20.2	930	23	AAS52391	E. coli DNA for ce
	24	185.6	18.0	2301	24	ABQ70706	Listeria monocytog
	25	162.4	17.6	319630	24	ABQ67194	Listeria innocua c
C	26	162.4	17.6	3011208	24	ABQ69245	Listeria innocua D
C	27	160.4	17.4	2944528	24	ABA03041	Listeria monocytog
	28	158	17.2	6021	20	AAV13395	Enterococcus faeca
	29	139.6	15.2	592	24	ABK78261	Bacillus clausii g
	30	136.6	14.8	46593	21	AAV81456	N. meningitidis pa
	31	136.6	14.8	34980	21	AAV21612	Neisseria meningit
	32	136.6	14.8	837096	21	AAV81489	N. meningitidis pa
	33	130.4	14.2	936	24	ABN92524	Staphylococcus epi
	34	128	13.9	936	23	AAS54395	Staphylococcus aur
	35	127.6	13.9	906	23	AAS51823	Staphylococcus aur
	36	126.4	13.7	927	19	AAV31440	Nucleotide sequenc
	37	126.4	13.7	999	22	AAV06204	Staphylococcus aur
	38	120.8	13.1	939	23	AAS54159	Pseudomonas aerugi
	39	118.4	12.9	1170	21	AAV51219	Staphylococcus aur
	40	116.8	12.7	672	22	AAV51330	S. epidermidis ope
C	41	116.8	12.7	4155	22	AAV54909	S. epidermidis gen
	42	112.6	12.2	954	24	ABQ90247	M. capsulatus gene
C	43	112	12.2	459	21	AAV21545	N. meningitidis pa
C	44	112	12.2	459	21	AAV81260	N. meningitidis pa
C	45	112	12.2	459	21	AAV53599	Neisseria meningit

## ALIGNMENTS

RESULT 1  
 AAV37239  
 ID AAV37239 standard; DNA; 921 BP.  
 XX  
 AC AAV37239;  
 XX  
 DT 24-SEP-1998 (first entry)  
 XX  
 DE DNA encoding a FabD polypeptide.  
 XX  
 KW FabD protein; malonyl-CoA:ACP family; diagnosis; infection; vaccine;  
 screen; ss.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 PH Key  
 CDS Location/Qualifiers  
 FT 1..921  
 FT /\*tag= a  
 FT /product= FabD  
 XX  
 PN WO9822133-A1.  
 XX  
 PD 28-MAY-1998.  
 XX  
 PF 14-NOV-1997; 97WO-US20992.  
 XX  
 PR Streptococcus pneu  
 XX  
 PR Streptococcus pneu  
 XX  
 PR Streptococcus pneu  
 XX  
 PR S. pneumoniae gene  
 XX  
 PR 2CFE 34 coding seq  
 XX  
 PR C-terminal coding  
 XX  
 PR DNA encoding a S.  
 XX  
 PR Streptococcus pneu

Gentry DR, Lonsdale JT, Payne DJ, Pearson SC, Van Aller G;

DR WPI; 1998-312173/27.  
 XX P-PSDB; AAW60856.  
 PT New isolated Streptococcus pneumoniae FapD gene - used to develop  
 PT products for the diagnosis, prevention and treatment of bacterial  
 PT diseases, particularly S. pneumoniae infection  
 XX  
 XX  
 PS Claim 6; Pages 5-6; 45pp; English.  
 XX  
 CC The present sequence encodes a FapD protein of Streptococcus pneumoniae  
 CC 0100993 (NCIMB 40800). The novel FapD polypeptides are related to other  
 CC proteins of the malonyl-CoA:ACP family. The products can be used for the  
 CC diagnosis of Streptococcus pneumoniae infections. Vectors containing  
 CC the FapD DNA sequence can be administered directly to a mammal to  
 CC produce the FapD peptide to provoke an antibody/T-cell response in  
 CC order to prevent a disease. The peptide can be used to screen for  
 CC compounds which modulate its activity.  
 XX  
 SQ Sequence 921 BP; 247 A; 193 C; 230 G; 251 T; 0 other;  
 Query Match 100.0%; Score 921; DB 19; Length 921;  
 Best Local Similarity 100.0%; P-Seq. No. 3.1e-281;  
 Matches 921; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATGACTAAACAGCCCTTTTATTTCTGCTCAAGGTGCCAGTATCTAGGGATGGGACGG 60  
 DB 1 ATGACTAAACAGCCCTTTTATTTCTGCTCAAGGTGCCAGTATCTAGGGATGGGACGG 60  
 QY 61 GATTCTATGATCAGTATCCGATTTGTCACAAAGAACGATGTGCGAGCGATGAGTGCCTC 120  
 DB 61 GATTCTATGATCAGTATCCGATTTGTCACAAAGAACGATGTGCGAGCGATGAGTGCCTC 120  
 QY 121 GGTATGATTTACGTTATCTCATCATACGGAAGAACAACTCAATCAGACCGGTAT 180  
 DB 121 GGTATGATTTACGTTATCTCATCATACGGAAGAACAACTCAATCAGACCGGTAT 180  
 QY 181 ACGCAACAGCCATCTAGGACCTTCGGTGTGCTATCTACGTTATTCACAAAGAAAGGC 240  
 DB 181 ACGCAACAGCCATCTAGGACCTTCGGTGTGCTATCTACGTTATTCACAAAGAAAGGC 240  
 QY 241 TATCAGCTGATATGGTGTGCTGTTGCTCTTGGAGAACTACTGCTTGGTGGCAAGC 300  
 DB 241 TATCAGCTGATATGGTGTGCTGTTGCTCTTGGAGAACTACTGCTTGGTGGCAAGC 300  
 QY 301 GCGCCTTTGGATTTTGAAGATCGGTTGCTTGTAGCTAAGCGTGGAGCCATATGGAA 360  
 DB 301 GCGCCTTTGGATTTTGAAGATCGGTTGCTTGTAGCTAAGCGTGGAGCCATATGGAA 360  
 QY 361 GAAGCGGCTCCTGCTGACTCTGGCAAGATGATGAGCAGTCTCAATACGCCAGTAGAGTC 420  
 DB 361 GAAGCGGCTCCTGCTGACTCTGGCAAGATGATGAGCAGTCTCAATACGCCAGTAGAGTC 420  
 QY 421 ATTGAAGAAGCCCTGTCAAAAAGCTTCTGAACCTGGAGTGGTGTACTCCAGGCCAACTATAAC 480  
 DB 421 ATTGAAGAAGCCCTGTCAAAAAGCTTCTGAACCTGGAGTGGTGTACTCCAGGCCAACTATAAC 480  
 QY 481 ACACCTGCACAAATCGTATCTGCGAGAGTGGTGTGATGATGATGATGATGATGATGATGAT 540  
 DB 481 ACACCTGCACAAATCGTATCTGCGAGAGTGGTGTGATGATGATGATGATGATGATGATGAT 540  
 QY 541 TTGCAAGAGCAGGTGCCAAAGCTTGTATCTCTTAAGGTGTGATGATGATGATGATGATGATGAT 600  
 DB 541 TTGCAAGAGCAGGTGCCAAAGCTTGTATCTCTTAAGGTGTGATGATGATGATGATGATGATGAT 600  
 QY 601 GCTCTCTTGGCCTGCTAGCCAGAACTAGCTGAACTCTAGCTGAGTAAAGTTTTCAT 660  
 DB 601 GCTCTCTTGGCCTGCTAGCCAGAACTAGCTGAACTCTAGCTGAGTAAAGTTTTCAT 660  
 QY 661 GATTCTTACTGTCCTAGTGGCAATACAGAGCTGCTGTGATGATGATGATGATGATGATGATGAT 720  
 DB 661 GATTCTTACTGTCCTAGTGGCAATACAGAGCTGCTGTGATGATGATGATGATGATGATGATGAT 720  
 QY 721 GCTCAGCTCTTGCAGCGTCAGGTCAAGGAACCCGTTGCTTCTATGAAAGTATTTGGGTC 780

Db 721 GCTCAGCTCTTGCAGCGTCAGGTCAAGGAACCCGTTGCTTCTATGAAAGTATTTGGGTC 780  
 QY 781 ATCAAGAGAGCAGGCATTAAGCAACTTTATCGAGATTGGACCGGGAAAGTCTTTGTGAGGT 840  
 Db 781 ATCAAGAGAGCAGGCATTAAGCAACTTTATCGAGATTGGACCGGGAAAGTCTTTGTGAGGT 840  
 QY 841 TTTGTTAAAAAATGATCAAACTGCTCACTTAGCTCATGTGGAAGATCAAGCGAGTTTA 900  
 Db 841 TTTGTTAAAAAATGATCAAACTGCTCACTTAGCTCATGTGGAAGATCAAGCGAGTTTA 900  
 QY 901 GTAGCACTTTTAGAAAAATAG 921  
 Db 901 GTAGCACTTTTAGAAAAATAG 921  
 RESULT 2  
 AAH90730  
 ID AAH90730 standard; DNA; 921 BP.  
 XX  
 AC AAH90730;  
 XX 02-OCT-2001 (first entry)  
 DT  
 XX CFE 34 coding sequence.  
 XX Antibacterial; vaccine; gene therapy; bacterial cell wall viability;  
 KW CFE; CEG; Conserved Essential Gene; bacterial infection;  
 KW antisense therapy; antibiotic resistance; ds.  
 XX Streptococcus pneumoniae.  
 XX WC200149721-A2.  
 PN 12-JUL-2001.  
 PD 29-DEC-2000; 2000WO-US35604.  
 XX 30-DEC-1999; 99US-0174089.  
 PR (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 PA Dougherty TJ, Pucci MJ, Dougherty BA, Davison DB, Bruccoleri RE;  
 PI Thanassi JA;  
 XX WPI; 2001-496721/54.  
 DR P-PSDB; AAM01031.  
 XX Nucleic acids encoding conserved essential genes involved in bacterial  
 PT replication which are potential targets for the treatment of antibiotic  
 PT resistant bacterial infections -  
 XX Claim 8; Pages 202-203; 380pp; English.  
 XX The present invention relates to nucleic acids (AAH90701-AAH90918)  
 CC encoding polypeptides (AAM01002-AAM01114), which are essential for the  
 CC viability of a bacterial cell wall. The acronym CFE stands for "CEG For  
 CC Expression", where CEG stands for "Conserved Essential Gene". The nucleic  
 CC acids are useful for detecting the presence of proteins essential for the  
 CC viability of a bacterial cell wall in samples such as cells, tissues,  
 CC biological fluids, blood, serum, nose, ear or throat swabs with ligands,  
 CC and for detecting corresponding target nucleic acid molecules with  
 CC complementary sequences. The nucleic acids are also useful for  
 CC determining whether a genomic nucleotide sequence of interest is  
 CC essential for viability of a bacterial cell or whether it resides within  
 CC an operon, by integrating an exogenous nucleotide sequence comprising a  
 CC portion of an open reading frame of the genomic sequence of interest  
 CC (comprising 200-500 base pairs) into the genomic sequence of interest  
 CC which confers a selectable phenotype to the cell, and determining cell  
 CC viability with a selection agent such as chloramphenicol. The nucleic  
 CC acids and proteins are also useful as vaccines and for treating bacterial  
 CC infections with gene therapy and antisense therapy. The nucleic acids  
 CC also enable identification of targets suitable for the treatment of

CC antibiotic resistant bacterial infections.

XX Sequence 921 BP; 246 A; 193 C; 232 G; 250 T; 0 other;

Query Match 99.1%; Score 913; DB 22; Length 921;  
Best Local Similarity 99.5%; Pred. No. 1.1e-278;  
Matches 916; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ATGACTAAACAGCCCTTTTATTTGCTGCTCAAGTGCCCAAGTATCTAGGATGGGACGG 60  
DB 1 ATGACTAAACAGCCCTTTTATTTGCTGCTCAAGTGCCCAAGTATCTAGGATGGGACGG 60  
QY 61 GATTTCTATGATCAGTATCCGATTCCTCAAGAAACGATTGATCGAGGAGTCAAGTGCCTC 120  
DB 61 GATTTCTATGATCAGTATCCGATTCCTCAAGAAACGATTGATCGAGGAGTCAAGTGCCTC 120  
QY 121 GGTATGATTTACGTTATCTCATCATGATGAGGAAAGCAAACTCAATCAAGCCGCTAT 180  
DB 121 GGTATGATTTACGTTATCTCATCATGATGAGGAAAGCAAACTCAATCAAGCCGCTAT 180  
QY 181 ACGCAACAGCCATCTAGCGACTTCGGTTCCTATCTACCGTTTATTGCAAGAAAGGCG 240  
DB 181 ACGCAACAGCCATCTAGCGACTTCGGTTCCTATCTACCGTTTATTGCAAGAAAGGCG 240  
QY 241 TATAGCCTGATATGCTGCTGGTTGCTCTTGAGAAATCTCTGCTTGGTGGCAAGC 300  
DB 241 TATAGCCTGATATGCTGCTGGTTGCTCTTGAGAAATCTCTGCTTGGTGGCAAGC 300  
QY 301 GCGCCTTGGATTTGAGATGCGGTTCCTTGGTGTAGTACGCTGAGGCTATATGAA 360  
DB 301 GCGCCTTGGATTTGAGATGCGGTTCCTTGGTGTAGTACGCTGAGGCTATATGAA 360  
QY 361 GAAGCGGCTCCTGCTGATCTGCGCAAGATGTTGAGTCTCAATAGCCAGTAGAGTC 420  
DB 361 GAAGCGGCTCCTGCTGATCTGCGCAAGATGTTGAGTCTCAATAGCCAGTAGAGTC 420  
QY 421 ATTGAAGAGCCTGTCAAAAGCTTCTGAACTTGGAGTGGTACTCCAGCCCAACTATAAC 480  
DB 421 ATTGAAGAGCCTGTCAAAAGCTTCTGAACTTGGAGTGGTACTCCAGCCCAACTATAAC 480  
QY 481 ACACCTGCACAAATCGTCAATGCTGGAGAGTGGTGGAGTGTATCGAGGGTTGAACCT 540  
DB 481 ACACCTGCACAAATCGTCAATGCTGGAGAGTGGTGGAGTGTATCGAGGGTTGAACCT 540  
QY 541 TTGCAAGAGCAGGTGCGCAACCGTTGATTCCTTTAAGTGTGAGTGCCTTTTCACAC 600  
DB 541 TTGCAAGAGCAGGTGCGCAACCGTTGATTCCTTTAAGTGTGAGTGCCTTTTCACAC 600  
QY 601 GCTCTCCTTGAGCCTGTAGCCAGAACTAGCTGAACTCTAGCTCAGGTAACTTTTCA 660  
DB 601 GCTCTCCTTGAGCCTGTAGCCAGAACTAGCTGAACTCTAGCTCAGGTAACTTTTCA 660  
QY 661 GATTTTACTTGTCCCTAGTCGCAATACAGAAGCTGCTGTGATGCAAAAGAGGACATT 720  
DB 661 GATTTTACTTGTCCCTAGTCGCAATACAGAAGCTGCTGTGATGCAAAAGAGGACATT 720  
QY 721 GCTAGCTCTTGAGCGTCAGGTCAAGAACCCGTTGTTCTATGAAAGTATGGGTC 780  
DB 721 GCTAGCTCTTGAGCGTCAGGTCAAGAACCCGTTGTTCTATGAAAGTATGGGTC 780  
QY 781 ATGCAAGAGCAGGCATGAAGCACTTTATCAGATTTGGACCGGGGAAAGTCTTGTCAAGT 840  
DB 781 ATGCAAGAGCAGGCATGAAGCACTTTATCAGATTTGGACCGGGGAAAGTCTTGTCAAGT 840  
QY 841 TTTGTTAAAAAATTGATCAAACTGCTCATTAGCTCATGTGGAAGATCAAGCGAGTTTA 900  
DB 841 TTTGTTAAAAAATTGATCAAACTGCTCATTAGCTCATGTGGAAGATCAAGCGAGTTTA 900  
QY 901 GTAGCATTCTTGAAGAAATAG 921  
DB 901 GTAGCATTCTTGAAGAAATAG 921

RESULT 3

AAV52140/c  
AAV52140 standard; DNA; 19702 BP.

XX AAV52140;  
AC AAV52140;

DT 23-OCT-1998 (first entry)

XX Streptococcus pneumoniae genome fragment SEQ ID NO:7.

DE Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;  
KW computer readable medium; vaccine; pharmaceutical composition; ds.

OS Streptococcus pneumoniae.

XX WO9818931-A2.

XX 07-MAY-1998.

PF 30-OCT-1997; 97WO-US19588.

PR 31-OCT-1996; 96US-0029960.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Barash SC, Choi GH, Dallon PJ, Dougherty BA, Fannon M;  
PI Kunsch CA, Rosen CA;

XX WPI; 1998-272225/24.

Computer-readable medium with recorded Streptococcus pneumoniae polynucleotide sequences - useful in diagnostic kits and assays, and pharmaceutical compositions and vaccines for Streptococcus pneumoniae

Claim 1; Page 194-205; 1409pp; English.

The present invention describes a computer readable medium which has the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524) recorded on it, or a representative fragment or a sequence at least 95% identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from Streptococcus pneumoniae. The present invention also describes an isolated nucleic acid molecule encoding a homologue of any of the fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced by a process comprising: (a) screening a genomic DNA library using as a probe a target sequence defined by any of the sequences in SEQ ID NO:1 to 391, identifying members of the library which contain sequences that hybridize to the target sequence and isolating the nucleic acid molecules from the members; or (b) isolating mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid molecules whose nucleotide sequence is homologous to amplification primers derived from the fragment of the S. pneumoniae genome to prime the amplification and isolating the amplified sequences. The computer readable medium can be used in a computer-based system for identifying fragments of the S. pneumoniae genome of commercial importance, or expression modulating fragments of the S. pneumoniae genome. Products from the present invention can be used in diagnosis kits and assays, and pharmaceutical compositions and vaccines for S. pneumoniae.

Sequence 19702 BP; 5736 A; 4548 C; 3728 G; 5690 T; 0 other;

Query Match 99.1%; Score 913; DB 19; Length 19702;

Best Local Similarity 99.5%; Pred. No. 5.8e-278;

Matches 916; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ATGACTAAACAGCCCTTTTATTTGCTGCTCAAGTGCCCAAGTATCTAGGATGGGACGG 60

DB 17149 ATGACTAAACAGCCCTTTTATTTGCTGCTCAAGTGCCCAAGTATCTAGGATGGGACGG 17090

QY 61 GATTTCTATGATCAGTATCCGATTTGTCATGAAGAACGATTCGAGCCAGTCGCTC 120

DB 17089 GATTTCTATGATCAGTATCCGATTTGTCATGAAGAACGATTCGAGCCAGTCGCTC 17030



QY 301 GCGCCCTTGGATTGTTGAAGATGCGGTGCGCTTGGTGTAGTAAGCGTGGAGCGTATATGGAA 360  
 Db 301 GCGCCCTTGGATTGTTGAAGATGCGGTGCGCTTGGTGTAGTAAGCGTGGAGCGTATATGGAA 360  
 QY 361 GAAGCGGCTCTGCTGACTCTGCGAAGATGCTAGCAGTCTCAATACGCCAGTAGAGGTC 420  
 Db 361 GAAGCGGCTCTGCTGACTCTGCGAAGATGCTAGCAGTCTCAATACGCCAGTAGAGGTC 420  
 QY 421 ATTGAAGAAGCGCTGTCAAAAAGCTTCTGAACCTTGGAGTGTACTCCAGCCAACTATPAAC 480  
 Db 421 ATTGAAGAAGCGCTGTCAAAAAGCTTCTGAACCTTGGAGTGTACTCCAGCCAACTATPAAC 480  
 QY 481 ACACCTGCACAAATCGTCAATTCCTGCGAAGTGTTCGAGTGTAGCAGCGGTGAACTT 540  
 Db 481 ACACCTGCACAAATCGTCAATTCCTGCGAAGTGTTCGAGTGTAGCAGCGGTGAACTT 540  
 QY 541 TTGCAAGAAGCAGGTGCCAAAGCTTCTGATTCCTCTTAAGTGTGTCAGGTCCCTTTCACACC 600  
 Db 541 TTGCAAGAAGCAGGTGCCAAAGCTTCTGATTCCTCTTAAGTGTGTCAGGTCCCTTTCACACC 600  
 QY 601 GCTCTCCTTGAAGCTGTAGCAGAACTAGCTGAAACTCTAGCTCAGTAAAGTTTTC 660  
 Db 601 GCTCTCCTTGAAGCTGTAGCAGAACTAGCTGAAACTCTAGCTCAGTAAAGTTTTC 660  
 QY 661 GATTTTACTTGTCCCTAGTCCGCAATACAGAACTGCTGTGATGCAAAAAGAGGACATT 720  
 Db 661 GATTTTACTTGTCCCTAGTCCGCAATACAGAACTGCTGTGATGCAAAAAGAGGACATT 720  
 QY 721 GCTCAGCTCTTGAAGCGGTGAGTCAAGAAACCGTTCGTTTCTATGAAAGTATTTGGGGTC 780  
 Db 721 GCTCAGCTCTTGAAGCGGTGAGTCAAGAAACCGTTCGTTTCTATGAAAGTATTTGGGGTC 780  
 QY 781 ATGCAAGAAGCAGGATGAAGCAACTTTATCGAGATTGGACCGGGGAAAGTCTTGTCAAGT 840  
 Db 781 ATGCAAGAAGCAGGATGAAGCAACTTTATCGAGATTGGACCGGGGAAAGTCTTGTCAAGT 840  
 QY 841 TTGTTTAAAAAATTGATCAAACTGCTCATCTAGCTCATGTGGAAGATCAAGCGAGTTTA 900  
 Db 841 TTGTTTAAAAAATTGATCAAACTGCTCATCTAGCTCATGTGGAAGATCAAGCGAGTTTA 900  
 QY 901 GTAGCATTGTAAGAAATAG 921  
 Db 901 GTAGCATTGTAAGAAATAG 921

RESULT 5  
 AAV65243  
 ID AAV65243 standard; DNA; 3580 BP.  
 XX AAV65243;  
 AC AAV65243;  
 XX XX  
 DT 24-DEC-1998 (first entry)  
 XX XX  
 DE S. pneumoniae gene fragment.  
 XX XX  
 KW Streptococcus pneumoniae protein; recombinant; gene expression;  
 KW DNA chip; virulence; antibody; infection; detection; treatment; ss.  
 XX XX  
 OS Streptococcus pneumoniae.  
 XX XX  
 PN W09826072-A1.  
 XX XX  
 PD 18-JUN-1998.  
 XX XX  
 PF 09-DEC-1997; 97NO-US22578.  
 XX XX  
 PR 13-DEC-1996; 96US-0036281.  
 XX XX  
 PA (ELIL) LILLY & CO ELI.

PI Balz RH, Burgett SG, Dehoff BS, Hoskins JA, Jaskunas BR;  
 PI Mills SJ, Norris FH, Peery RB, Rokey PK, Rostek PR;  
 PI Skatrud PL, Smith MC, Solenberg PJ, Treadway PJ;

PI Young Bellido ML;  
 XX DR WPI; 1998-348529/30.  
 XX PT Streptococcus pneumoniae nucleic acid sequences - used in DNA chips  
 PT for evaluating gene expression, and identification of virulence  
 PT genes  
 XX Claim 1; Pages 96-98; 33pp; English.  
 XX PS  
 CC This sequence is a gene fragment from the Streptococcus pneumoniae  
 CC genome. The invention provides DNA sequences (AAV65201 to AAV65304) from  
 CC the Streptococcus pneumoniae genome and corresponding protein sequences  
 CC (AAV6505 to AAV6529). A recombinant host containing a vector comprising  
 CC any of the above nucleic acids can be used for the recombinant expression  
 CC of the protein sequences. The invention also provides a DNA chip having  
 CC arrayed on it at least 15 base pair fragment of any one or more of these  
 CC DNA sequences. The DNA chip can be used methods for evaluating gene  
 CC expression in S. pneumoniae and for identifying virulence genes in  
 CC S. pneumoniae. Antibodies that selectively bind to the above proteins or  
 CC peptide fragments can be used to treat S. pneumoniae infection. The  
 CC antibodies can also be used to detect S. pneumoniae cells.  
 XX Sequence 3580 BP; 1035 A; 666 C; 894 G; 985 T; 0 other;

Query Match 99.0%; Score 911.4; DB 19; Length 3580;  
 Best Local Similarity 99.3%; Pred. No. 7.3e-278; Indels 0; Gaps 0;  
 Matches 915; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 QY 1 ATGACTAAACAGCCCTTTTATTTCTGCTCAAGTGCCTGATCTAGGATGGAGCG 60  
 Db 718 ATGACTAAACAGCCCTTTTATTTCTGCTCAAGTGCCTGATCTAGGATGGAGCG 777  
 QY 61 GATTTCTATGATCAGTATCCGATTCGAAAGAAAGATTCGAGCGAGTCAGGTGTC 120  
 Db 778 GATTTCTATGATCAGTATCCGATTCGAAAGAAAGATTCGAGCGAGTCAGGTGTC 837  
 QY 121 GGTATGATTTACGTTATCTCATCGATACGGAAGAAAGCAAACTCAATCAGACCCGCTAT 180  
 Db 838 GGTATGATTTGCGTTATCTCATCGATACGGAAGAAAGCAAACTCAATCAGACCCGCTAT 897  
 QY 181 AGCAACAGCCATCTAGCGACTTCGGTTGCTATCTACCGTTTATTCGAGAAAGGCG 240  
 Db 898 AGCAACAGCCATCTAGCGACTTCGGTTGCTATCTACCGTTTATTCGAGAAAGGCG 957  
 QY 241 TATCAGCTGATGTTGCTGTTTGTCTCTGGAGATGCTGAGTCTCAATACGCCAGTAGAGGTC 300  
 Db 958 TATCAGCTGATGTTGCTGTTTGTCTCTGGAGATGCTGAGTCTCAATACGCCAGTAGAGGTC 1017  
 QY 301 GCGGCTTGGATTTTGAAGATCGGTTCCTTGGTAGCTAGCGTGGAGCCCTATATGGAA 360  
 Db 1018 GCGGCTTGGATTTTGAAGATCGGTTCCTTGGTAGCTAGCGTGGAGCCCTATATGGAA 1077  
 QY 361 GAAGCGGCTCCTGCTGACTCTGCGAAGATGGTAGCAGTCTCAATACGCCAGTAGAGGTC 420  
 Db 1078 GAAGCGGCTCCTGCTGACTCTGCGAAGATGGTAGCAGTCTCAATACGCCAGTAGAGGTC 1137  
 QY 421 ATTGAAGAAGCGCTGTCAAAAAGCTTCTGAACTTGGAGTGTGTTACTCCAGCCAACTATAAC 480  
 Db 1138 ATTGAAGAAGCGCTGTCAAAAAGCTTCTGAACTTGGAGTGTGTTACTCCAGCCAACTATAAC 1197  
 QY 481 ACACCTGCACAAATCGTCAATTCCTGCGAAGTGTTCGAGTGTAGCAGCGGTGAACTT 540  
 Db 1198 ACACCTGCACAAATCGTCAATTCCTGCGAAGTGTTCGAGTGTAGCAGCGGTGAACTT 1257  
 QY 541 TTGCAAGAAGCAGGTGCCAAAGCTTGTATTCCTCTTAAGGTGTAGGTCCCTTTCACACC 600  
 Db 1258 TTGCAAGAAGCAGGTGCCAAAGCTTGTATTCCTCTTAAGGTGTAGGTCCCTTTCACACC 1317  
 QY 601 GCTCTCCTTGAAGCTGTAGCAGAACTAGCTGAAACTCTAGCTCAGTAAAGTTTTC 660  
 Db 1318 TCTCTCCTTGAAGCTGTAGCAGAACTAGCTGAAACTCTAGCTCAGTAAAGTTTTC 1377



Db 901 GTAGCATTTTAGAAAA 918  
 RESULT 7  
 AAV37241  
 ID AAV37241 standard; DNA; 547 BP.  
 XX  
 AC AAV37241;  
 DT 24-SEP-1998 (first entry)  
 DE C-terminal coding region encoding a FabD polypeptide embodiment.  
 XX  
 KW FabD protein; malonyl-CoA:ACP family; diagnosis; infection; vaccine;  
 screen; ss.  
 OS Streptococcus pneumoniae.  
 XX  
 PN WO9822133-A1.  
 XX  
 PD 28-MAY-1998.  
 XX  
 PF 14-NOV-1997; 97WO-US20992.  
 XX  
 PR 18-NOV-1996; 95US-0031160.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 XX  
 PI Gentry DR, Lonsdale JT, Payne DJ, Pearson SC, Van Aller G;  
 DR WPI; 1998-312173/27.  
 DR P-PSDB; AAW60858.  
 XX  
 PT New isolated Streptococcus pneumoniae FabD gene - used to develop  
 PT products for the diagnosis, prevention and treatment of bacterial  
 PT diseases, particularly S. pneumoniae infection  
 PS Disclosure; Pages 6-7; 45pp; English.  
 XX  
 CC The present sequence represents the C-terminal coding region of FabD  
 CC embodiment (coding begins at nucleotide 2) of Streptococcus pneumoniae  
 CC 010093 (NCIMB 40800). The novel FabD polypeptides are related to other  
 CC proteins of the malonyl-CoA:ACP family. The products can be used for the  
 CC diagnosis of Streptococcus pneumoniae infections. Vectors containing  
 CC the FabD DNA sequence can be administered directly to a mammal to  
 CC produce the FabD peptide to provoke an antibody/i-cell response in  
 CC order to prevent a disease. The peptide can be used to screen for  
 CC compounds which modulate its activity.  
 XX  
 SQ Sequence 547 BP; 154 A; 115 C; 132 G; 146 T; 0 other;  
 Query Match 57.5%; Score 529.4; DB 19; Length 547;  
 Best Local Similarity 98.0%; Pred. No. 3.2e-157;  
 Matches 536; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
 QY 372 TGCTGACTCTGCGAAGATGTTGAGAGTTCTCAATACGCCAGTAGAGTTCATTGAAGAAGC 431  
 Db 1 TGTGTAAGTTGCGAAGATGTTGAGAGTTCTCAATACGCCAGTAGAGTTCATTGAAGAAGC 60  
 QY 432 CTGTCAAAAAGCTTCTGAACTTGGAGTGGTGTACTCCAGCCCACTATACACACCTGACCA 491  
 Db 61 CTGTCAAAAAGCTTCTGAACTTGGAGTGGTGTACTCCAGCCCACTATACACACCTGACCA 120  
 QY 492 AATGCTCATTCTGGAGAAGTGGTTGCAAGTTGATCGAGCGGTTGAAGTTTGAAGAAGC 551  
 Db 121 AATGCTCATTCTGGAGAAGTGGTTGCAAGTTGATCGAGCGGTTGAAGTTTGAAGAAGC 180  
 QY 552 AGGTGCCAAAGCTTGTATCTCTTAAGGTGTACAGGTCCCTTTACACCGCTCTCCTTGA 611  
 Db 181 AGGTGCCAAAGCTTGTATCTCTTAAGGTGTACAGGTCCCTTTACACCGCTCTCCTTGA 240  
 QY 612 GCCTGCTAGCCAGAAAGCTAGCTGAACTCTAGCTCAGGTAAGTTTTCAGATTTTACTTG 671

Db 241 GCCTGCTAGCCAGAACTAGCTGAACCTAGCTAGTAACTTTTTCAGATTTTACTTG 300  
 QY 672 TCCCTAGTCGGCAATACAGAAGCTGCTGTATGCAAAAAGAGACATTGCTCAGCTCTT 731  
 Db 301 TCCCTAGTCGGCAATACAGAAGCTGCTGTATGCAAAAAGAGACATTGCTCAGCTCTT 360  
 QY 732 GACGCGTCAGGTCAAGGAACCCGTTTCGTTTCTATGAAAGTATTGGGGTCATGCAAGAAGC 791  
 Db 361 GACGCGTCAGGTCAAGGAACCCGTTTCGTTTCTATGAAAGTATTGGGGTCATGCAAGAAGC 420  
 QY 792 AGCATATAGCAAACTTTATCGAGATTGGACCGGGGAAAGTCTTGTGAGGTTTGTAAAAA 851  
 Db 421 AGCATATAGCAAACTTTATCGAGATTGGACCGGGGAAAGTCTTGTGAGGTTTGTAAAAA 480  
 QY 852 AATTGATCAAACTGCTCACTTAGCTCATGTGAAAGTCAAGCGAGTTTAGTAGCACTTTT 911  
 Db 481 AATTGATCAAACTGCTCACTTAGCTCATGTGAAAGTCAAGCGAGTTTAGTAGCACTTTT 540  
 QY 912 AGAAAAA 918  
 Db 541 AGAAAAA 547  
 RESULT 8  
 AAT98542  
 ID AAT98542 standard; DNA; 1196 BP.  
 XX  
 AC AAT98542;  
 DT 06-NOV-1998 (first entry)  
 DE DNA encoding a S. pneumoniae 3-oxoacyl reductase precursor protein.  
 XX  
 KW Streptococcus pneumoniae protein; genetic immunisation; antagonist;  
 KW immunological response; inoculation; antibody production; inhibitor;  
 KW T cell immune response; antimicrobial compound; bacterial adhesion;  
 KW extracellular matrix protein; protein-mediated cell invasion; wound;  
 KW pathogenesis; ss.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 586..1146  
 FT /\*tag= a  
 XX  
 PN WO9743303-A1.  
 XX  
 PD 20-NOV-1997.  
 XX  
 PF 14-MAY-1997; 97WO-US07950.  
 XX  
 PR 14-MAY-1996; 95US-0017670.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 XX  
 PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO;  
 PI Stodola RK;  
 XX  
 DR WPI; 1998-008793/01.  
 DR P-PSDB; AAW38474.  
 XX  
 PT Novel Streptococcus pneumoniae proteins and related DNA - useful for  
 PT diagnosing anti-microbial agents for treatment of bacterial  
 PT infections  
 XX  
 PS Claim 4; Page 90; 483pp; English.  
 XX  
 CC This sequence encodes a Streptococcus pneumoniae protein that (based on  
 CC homology with a Cuphea lanceolata protein) is a 3-oxoacyl reductase  
 CC precursor protein, and represents a DNA sequence of the invention.  
 CC The DNA sequences were isolated from Streptococcus pneumoniae strain

CC 0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the  
 CC invention can be used to identify compounds which interact with and  
 CC inhibit or activate the activity of the proteins. Antagonists can be  
 CC used to treat diseases caused by S. pneumoniae proteins, through genetic  
 CC immunisation. They can also be used to induce an immunological response  
 CC in a mammal by inoculation with the S. pneumoniae proteins or delivery  
 CC of the encoding nucleic acids in a vector adequate to produce antibody  
 CC and/or T cell immune responses to protect the animal from disease. The  
 CC proteins can also be used to identify antimicrobial compounds which are  
 CC capable of inhibiting their bioactivity. In particular the proteins of  
 CC the invention can be used to prevent adhesion of bacteria to mammalian  
 CC extracellular matrix proteins on in-dwelling devices or in wounds, to  
 CC block protein-mediated mammalian cell invasion, and to block the normal  
 CC progression of pathogenesis in infections initiated other than by the  
 CC implantation of in-dwelling devices or other surgical techniques.  
 XX  
 XX Sequence 1199 BP; 355 A; 223 C; 286 G; 332 T; 0 other;

Query Match 56.8%; Score 523.2; DB 19; Length 1196;  
 Best Local Similarity 99.1%; Pred. No. 4.6e-155;  
 Matches 547; Conservative 0; Mismatches 3; Indels 2; Gaps 2;  
 QY 372 TGTGACTCTGCAAGATGGTAGCAGTTCTCAATAGCGCAGTAGAGGTCAATTGAAGAAGC 431  
 Db 1 TGTGAACTTGGCAAGATGGTAGCAGTTCTCAATAGCGCAGTAGAGGTCAATTGAAGAAGC 60  
 QY 432 CTGTCGCAAGAGCTTCT-GAAGTGGAGTGGTTACTCCAGCCCACTATACACACCTGCAC 490  
 Db 61 CTGTCGCAAGAGCTTCTGGAACTTGGAGTGGTTACTCCAGCCCACTATACACACCTGCAC 120  
 QY 491 AAATCGTCAATTCGTGAGAGTGGTTGCGATTGATCGAGCGGTGAACTTTTGAAGAAG 550  
 Db 121 AAATCGTCAATTCGTGAGAGTGGTTGCGATTGATCGAGCGGTGAACTTTTGAAGAAG 180  
 QY 551 CAGTCCCAAGCGTTGATTCCTCTTAAGTGTGAGTGGTCCCTTTACACCGCTCTCCTTG 610  
 Db 181 CAGTCCCAAGCGTTGATTCCTCTTAAGTGTGAGTGGTCCCTTTACACCGCTCTCCTTG 240  
 QY 611 AGCTCTGCTAGCCAGAACTAGTGAAGTCTAGCTCAGTAAAGTCTTTTTCAGATTTTACTT 670  
 Db 241 AGCTCTGCTAGCCAGAACTAGTGAAGTCTAGCTCAGTAAAGTCTTTTTCAGATTTTACTT 300  
 QY 671 GTCCCTTAGTCGCAATCAGAGAGTGTGTGATGCAAAAGAGACATTTGCTAGCTCT 730  
 Db 301 GTCCCTTAGTCGCAATCAGAGAGTGTGTGATGCAAAAGAGACATTTGCTAGCTCT 360  
 QY 731 TGACGGGTGAGGTCAGGCAAGCAACCGTTCGTTTCTATGAAAGTATTGGGGTTCATGCAAGAAG 790  
 Db 361 TGACGGGTGAGGTCAGGCAAGCAACCGTTCGTTTCTATGAAAGTATTGGGGTTCATGCAAGAAG 420  
 QY 791 CAGGCAATAGCAACTTTTA-TCGAGATTGCAACGGGGAAGCTTTGTCAGGTTTGTGTTAA 849  
 Db 421 CAGGCAATAGCAACTTTTATTCGAGATTGCAACGGGGAAGCTTTGTCAGGTTTGTGTTAA 480  
 QY 850 AAAATTGATCAACTCTCCTCACTTAGCTCATGTGGAAGTCAAGCGAGTTTGTAGCACTT 909  
 Db 481 AAAATTGATCAACTCTCCTCACTTAGCTCATGTGGAAGTCAAGCGAGTTTGTAGCACTT 540  
 QY 910 TTAGAAAAATAG 921  
 Db 541 TTAGAAAAATAG 552

RESULT 9  
 AAX30801/c  
 ID AAX30801 standard; DNA; 1199 BP.  
 XX  
 XX AAX30801;  
 XX  
 XX 20-MAY-1999 (first entry)  
 XX  
 XX Streptococcus pneumoniae genomic DNA sequence SEQ ID NO:78.  
 XX

KW Streptococcus pneumoniae strain 0100993; vaccine; immune response;  
 KW streptococcal infection; pneumococcal; ss.  
 XX  
 OS Streptococcus pneumoniae.  
 PN W09737026-A1.  
 PD 09-OCT-1997.  
 XX  
 XX 01-APR-1997; 97WO-US05306.  
 PR 22-AUG-1996; 96US-0025788.  
 PR 02-APR-1996; 96US-0014690.  
 XX  
 XX (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 XX  
 XX Black MT, Hodgson JE, Knowles DJC, Nicholas RO;  
 PI Stodola RK;  
 XX  
 XX WPI; 1997-503111/46.  
 DR P-PSDB; AAY11212.  
 XX  
 XX Nucleic acids encoding pneumococcal polypeptide(s) - useful in  
 PT vaccines, drug screening, etc  
 PT  
 XX  
 PS Claim 5; Page 105-106; 354pp; English.  
 XX  
 CC AAX30724 to AAX30946 represent genomic DNA sequences isolated from  
 CC Streptococcus pneumoniae strain 0100993. These genomic DNA sequences  
 CC encode the novel proteins given in AAY1114 to AAY11367. The proteins,  
 CC isolated from Streptococcus pneumoniae, can be used in vaccines against  
 CC streptococcal infections and in assays for identifying compounds that  
 CC inhibit or activate the activity of the proteins. The antagonists can  
 CC be used to treat an individual having need to inhibit a bacterial  
 CC protein. Vectors expressing the proteins can be used to induce a  
 CC protective immune response in mammals.  
 XX  
 SQ Sequence 1199 BP; 334 A; 287 C; 223 G; 355 T; 0 other;  
 Query Match 56.8%; Score 523.2; DB 18; Length 1199;  
 Best Local Similarity 99.1%; Pred. No. 4.6e-155;  
 Matches 547; Conservative 0; Mismatches 3; Indels 2; Gaps 2;  
 QY 372 TGTGACTCTGCAAGATGGTAGCAGTTCTCAATAGCGCAGTAGAGGTCAATTGAAGAAGC 431  
 Db 1199 TGTGAACTTGGCAAGATGGTAGCAGTTCTCAATAGCGCAGTAGAGGTCAATTGAAGAAGC 1140  
 QY 432 CTGTCGCAAGAGCTTCT-GAAGTGGAGTGGTTACTCCAGCCCACTATACACACCTGCAC 490  
 Db 1139 CTGTCGCAAGAGCTTCTGGAACTTGGAGTGGTTACTCCAGCCCACTATACACACCTGCAC 1080  
 QY 491 AAATCGTCAATTCGTGAGAGTGGTTGCGATTGATCGAGCGGTGAACTTTTGAAGAAG 550  
 Db 1079 AAATCGTCAATTCGTGAGAGTGGTTGCGATTGATCGAGCGGTGAACTTTTGAAGAAG 1020  
 QY 551 CAGTCCCAAGCGTTGATTCCTCTTAAGTGTGAGTGGTCCCTTTACACCGCTCTCCTTG 610  
 Db 1019 CAGTCCCAAGCGTTGATTCCTCTTAAGTGTGAGTGGTCCCTTTACACCGCTCTCCTTG 960  
 QY 611 AGCTCTGCTAGCCAGAACTAGTGAAGTCTAGCTCAGTAAAGTCTTTTTCAGATTTTACTT 670  
 Db 959 AGCTCTGCTAGCCAGAACTAGTGAAGTCTAGCTCAGTAAAGTCTTTTTCAGATTTTACTT 900  
 QY 671 GTCCCTTAGTCGCAATCAGAGAGTGTGTGATGCAAAAGAGGACATTTGCTAGCTCT 730  
 Db 899 GTCCCTTAGTCGCAATCAGAGAGTGTGTGATGCAAAAGAGGACATTTGCTAGCTCT 840  
 QY 731 TGACGGGTGAGGTCAGGCAAGCAACCGTTCGTTTCTATGAAAGTATTGGGGTTCATGCAAGAAG 790  
 Db 839 TGACGGGTGAGGTCAGGCAAGCAACCGTTCGTTTCTATGAAAGTATTGGGGTTCATGCAAGAAG 780  
 QY 791 CAGGCAATAGCAACTTTTA-TCGAGATTGCAACGGGGAAGCTTTGTCAGGTTTGTGTTAA 849

779 CAGCATAGCACTTTATTCGAGATTGGACCGGGGAAAGCTCTGTCAGGTTTGTAAA 720  
 850 AAAATTGATCAAACTGCTCACTTAGCTCATGTGGAAGATCAACGAGATTAGTACACTT 909  
 719 AAAATTGATCAAACTGCTCACTTAGCTCATGTGGAAGATCAACGAGATTAGTACACTT 660  
 910 TTAGAAAAAATAG 921  
 659 TTAGAAAAAATAG 648

## RESULT 10

AAV37406/C

ID AAV37406 standard; DNA; 1209 BP.

AC AAV37406;

XX 13-OCT-1998 (first entry)

XX Streptococcus pneumoniae coding region.

XX coding region; ORF; open reading frame; antibacterial;

XX infection; prevention; meningitis; ss.

XX Streptococcus pneumoniae.

XX Key Location/Qualifiers

XX CDS complement (731..1123)

XX /\*tag= a

XX /product= putative malonyl coenzyme A-acyl carrier

XX protein transacylase homologue

XX WO9813689-A1.

XX 14-MAY-1998.

XX 27-OCT-1997; 97WO-US19226.

XX 01-NOV-1996; 96US-0029930.

XX (SMIK ) SMITHKLINE BEECHAM CORP.

XX (SMIK ) SMITHKLINE BEECHAM PLC.

XX Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;

XX Reid RH, Zarfos PN;

XX WPI; 1998-286586/25.

XX P-PSDB; AAW61018.

XX New isolated nucleic acids from Streptococcus pneumoniae - useful,

XX e.g. for identifying anti-bacterial(s) for treatment and prevention

XX of meningitis

XX Claim 1; Page 111-112; 130pp; English.

XX The sequence is that of a coding region isolated from

XX S. pneumoniae. Its encoded protein, or agonists of it,

XX may be useful as an antibacterial for treatment or

XX prevention of infection, specifically caused by S. pneumoniae

XX (particularly meningitis) but possibly also Helicobacter

XX pylori (ulcers and gastric cancer). It may be of particular

XX use before insertion of an in-dwelling device or any other

XX invasive procedure. The protein, or nucleic acid encoding

XX it, can also be used in vaccines to induce a cellular

XX and/or humoral immune response, or to screen for other

XX antibacterials. The DNA may also contain flanking sequences

XX that are potential sources of control elements for bacterial

XX gene expression. Detecting a sequence encoding the protein

XX can be used diagnostically, e.g. to detect a mutation for

XX serotyping or classifying infectious agents.

XX Sequence 1209 BP; 336 A; 288 C; 227 G; 358 T; 0 other;

XX (CHIR-) CHIRON SPA.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Query Match 56.8%; Score 523.2; DB 19; Length 1209;

Best Local Similarity 99.1%; Pred. No. 4.6e-155;

Matches 547; Conservative 0; Mismatches 3; Indels 2;

Gaps 2;

372 TGCTGACTCTGGCAAGATGGTAGGAGTTCTCAATACGCCAGTAGAGGTCAATTGAAGAAGC 431

1209 TGCTGAACTTGGCAAGATGGTAGGAGTTCTCAATACGCCAGTAGAGGTCAATTGAAGAAGC 1150

432 CTGTCAAAAAGCTTCT-GAACTTGGAGTGGTTACTCCAGCCAACTATAACACACTGAC 490

1149 CTGTCAAAAAGCTTCTGGAACCTTGGAGTGGTTACTCCAGCCAACTATAACACACTGAC 1090

491 AAATCGTCAATTCGTGGAGAGTGGTTGCAGTTGATCGAGCGGTTGAACCTTTTGAAGAAG 550

1089 AAATCGTCAATTCGTGGAGAGTGGTTGCAGTTGATCGAGCGGTTGAACCTTTTGAAGAAG 1030

551 CAGGTGCCAAAAGCTTGGTTGATTCCTCTTAAGGTGTGAGTCCCTTTTCAACCCGCTCTCCCTTG 610

1029 CAGGTGCCAAAAGCTTGGTTGATTCCTCTTAAGGTGTGAGTCCCTTTTCAACCCGCTCTCCCTTG 970

611 AGCTGTAGTGCAGAACTAGCTGAACTCTAGCTCAGGTAAAGTTTTCAGATTTTACTT 670

989 AGCTGTAGTGCAGAACTAGCTGAACTCTAGCTCAGGTAAAGTTTTCAGATTTTACTT 910

671 GTCCCTAGTGGCAATACAGAACTGCTGTGATGCAAAAAGAGAGACATTGCTCAGCTCT 730

909 GTCCCTAGTGGCAATACAGAACTGCTGTGATGCAAAAAGAGAGACATTGCTCAGCTCT 850

731 TGACGCTCAGTCAAGGAACCCGTTCTGTTCTATGAAGTATTGGGGTCAATGAAGAAG 790

849 TGACGCTCAGTCAAGGAACCCGTTCTGTTCTATGAAGTATTGGGGTCAATGAAGAAG 790

791 CAGGCATAAGCAACTTTA-TGAGATTGGACCGGGGAAAGTCTTGTGAGGTTTGTGTTAAA 849

789 CAGGCATAAGCAACTTTATTCGAGTTGGACCGGGGAAAGTCTTGTGAGGTTTGTGTTAAA 730

850 AAAATTGATCAAACTGCTCACTTAGCTCATGTGGAAGATCAAGCGAGTTTAGTAGCACTT 909

729 AAAATTGATCAAACTGCTCACTTAGCTCATGTGGAAGATCAAGCGAGTTTAGTAGCACTT 670

910 TTAGAAAAAATAG 921

659 TTAGAAAAAATAG 658

RESULT 11

ID AAV37406 standard; DNA; 942 BP.

XX AAV37406;

XX 01-JUL-2002 (first entry)

XX Streptococcus pneumoniae coding region.

XX coding region; ORF; open reading frame; antibacterial;

XX infection; prevention; meningitis; ss.

XX Streptococcus pneumoniae.

XX Key Location/Qualifiers

XX CDS complement (731..1123)

XX /\*tag= a

XX /product= putative malonyl coenzyme A-acyl carrier

XX protein transacylase homologue

XX WO9813689-A1.

XX 14-MAY-1998.

XX 27-OCT-1997; 97WO-US19226.

XX 01-NOV-1996; 96US-0029930.

XX (SMIK ) SMITHKLINE BEECHAM CORP.

XX (SMIK ) SMITHKLINE BEECHAM PLC.

XX Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;

XX Reid RH, Zarfos PN;

XX WPI; 1998-286586/25.

XX P-PSDB; AAW61018.

XX New isolated nucleic acids from Streptococcus pneumoniae - useful,

XX e.g. for identifying anti-bacterial(s) for treatment and prevention

XX of meningitis

XX Claim 1; Page 111-112; 130pp; English.

XX The sequence is that of a coding region isolated from

XX S. pneumoniae. Its encoded protein, or agonists of it,

XX may be useful as an antibacterial for treatment or

XX prevention of infection, specifically caused by S. pneumoniae

XX (particularly meningitis) but possibly also Helicobacter

XX pylori (ulcers and gastric cancer). It may be of particular

XX use before insertion of an in-dwelling device or any other

XX invasive procedure. The protein, or nucleic acid encoding

XX it, can also be used in vaccines to induce a cellular

XX and/or humoral immune response, or to screen for other

XX antibacterials. The DNA may also contain flanking sequences

XX that are potential sources of control elements for bacterial

XX gene expression. Detecting a sequence encoding the protein

XX can be used diagnostically, e.g. to detect a mutation for

XX serotyping or classifying infectious agents.

XX Sequence 1209 BP; 336 A; 288 C; 227 G; 358 T; 0 other;

XX (CHIR-) CHIRON SPA.

XX

XX

XX

XX

XX

XX

XX

(GENO-) INST GENOMIC RES.

PA Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;  
PI Tettelin H;XX  
DR WPI; 2002-352536/38.  
DR P-PSDB; ABP28010.XX New Streptococcus protein for the treatment or prevention of infection  
PT or disease caused by Streptococcus bacteria, such as meningitis, and  
PI for detecting a compound that binds to the protein -XX  
PS Claim 7; Page 3863; 4525pp; English.XX The invention relates to a protein (ABP25413-ABP30895) from group B  
CC Streptococcus/SBS (Streptococcus agalactiae) or group A streptococcus/GAS  
CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in  
CC the specification. The proteins have antibacterial and anti-inflammatory  
CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and  
CC antibodies that bind (I) are used in the manufacture of medicaments for  
CC the treatment or prevention of infection or disease caused by  
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
CC biological sample. (I) is used to determine whether a compound binds to  
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
CC used as a vaccine or diagnostic composition. The disease caused by  
CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
CC acid encoding (I) may be used to recombinantly produce (I) and may be  
CC used in gene therapy. Antibodies to (I) are used for affinity  
CC chromatography, immunoassays, and distinguishing/identifying  
CC Streptococcus proteins.XX  
SQ Sequence 942 BP; 285 A; 159 C; 217 G; 281 T; 0 other;

Query Match 47.7%; Score 439.4; DB 24; Length 942;

Best Local Similarity 67.3%; Pred. No. 1.5e-128;

Matches 620; Conservative 0; Mismatches 301; Indels 0; Gaps 0;

QY 1 ATGACTAAACAGCCTTTTATTTGCTGTCAGTGCCAGTATCTAGGGATGGGCGG 60  
 Db 7 ATGACAAAGACAGCCTTTTATTTGCGGTCAAGTGCTCAAAAATTAGGGATGGCAAGG 66  
 QY 61 GATTTCATGATCAGTATCGATTGTCAAGAAACGATTGATCGAGCGAGTCAGGTGCTC 120  
 Db 67 GATTTTATGATAACTTTGCTATTGTGAAGAAACCTTTGATCAAGCTAGTCAATATG 126  
 QY 121 GGTATGATTAGTTAGTTATCTCATCGATACGGAAGAGACAACTCAATCAGACCCGCTAT 180  
 Db 127 GGATACGATTGCGCGCTTGATTGATAGTACGAGTTAAACCTTAATCAAACTAGCTAT 186  
 QY 181 ACCCAACCCAGCCTTCTAGCGCTTCGTTGCTATCTACCGTTTATTGCAAGAAAGGCG 240  
 Db 187 ACCCAACCCAGCTATTTGACATCATCAATGCTATTTACCGTGTGTTGGGCTTACATCAC 246  
 QY 241 TATCAGCCTGATATGTTGCTGTTTGTCTCTTGGAGAAATCTCTGCCTTGGTGGCAAGC 300  
 Db 247 GTTAAACCGGATATGTTAGTGTGCTCTTCTTAGGAGAAATCTCAGCTTTGGTAGCATCA 306  
 QY 301 GCGGCTTGATTTTGAAGATCGGTTGCTTGGTAGCTAAGCGTGGAGCCTATATGAA 360  
 Db 307 GGGGCACTCTCTTTTGAAGATACCCCTATCTTAGTAGTAAGAGAGGCGGCTTGTAGGAG 366  
 QY 361 GAAGCGGCTCTCTGCTGACTCTGGCAAGATGGTAGAGTTCTCAATACGCCAGTAGAGTTC 420  
 Db 367 GAAGCAGCACCAAGGATCTGAAAGATGGTTGCGGTTATGAATACAGACGTGCAAGTC 426  
 QY 421 ATTGAAGAACCTTGTCAAAAAGCTTCTGAACITGGAGTGGTTACTCCAGCCAACTATAAC 480  
 Db 427 ATGCAAGAAAGTCTGTCTCAAAATAGTCTGTAACATAGGAGTGGTTGCTCCAGCAAACTATAT 486  
 QY 481 ACACCTTGCAAAATCGTCATTGCTGAGAGAGTGGTTGTCAGTTGATCGAGCGGTGCACTT 540  
 Db 487 ACTCCTTAGTCAAAATTGTATTTGCTGTCAGACAGATGCTGTGAAAGTGGCAGTTGACTT 546

QY 541 TTGCAAGACGAGGTGCCAAACCGCTTGATTCCTCTTAAGGTGTCAGGTCCCTTTACACC 600  
 Db 547 TTAAGGAAAGGGGAGTTAAGCGCTTTAATCCCTTTAAACGCTGTACGTCCTTTCCACACT 606  
 QY 601 GCTCTCCTTGAGCTGCTAGCCAGAAACTAGCTGAAACTCTAGCTCAGGTAAAGTTTTCATCA 660  
 Db 607 GCITTTGTAGAACGAGTAGACCGCTGTTGGCTAAAGAGTTGAAAGATACAACTTCAGT 666  
 QY 661 GATTTACTTGTCCCTAGTCGGCAATACAGAAGCTGCTGATGATGCAAAAAGAGGACATT 720  
 Db 667 GACTTCAAGATTCTCTTTGGTTGGTAATACCGAAGCTAATATTATGGAAGAACCCGTATC 726  
 QY 721 GCTCAGCTCTTGACGGCTCAGGTCAAGAACCGCTTCGTTCTTCTATCAAGATATTGGGTC 780  
 Db 727 CCAGAACTATTAGCCGCTCAAGTCATGGAGCCCTGTCGTTTATGACAGTGTGGGACT 786  
 QY 781 ATGCAAGACGAGCGCATTAAGCAACTTATGAGATTGGACCGGGGAAAGTCTTTGTCAGGT 840  
 Db 787 TTAGTAGAAAGTGGCATTAACACAATTCATTGAGGTAGGACCAAGTAAAGTTTTGACAGGT 846  
 QY 841 TTTGTTAAAAAATTGATCAAACTGCTCACTTAGCTCATGTGGAAGATCAAGCGAGTTTA 900  
 Db 847 TTTGTGAAGAAATTGATAAAAAATTTACTATGTAAGTGTGTAAGAAACATGGTAAGCTTA 906  
 QY 901 GTAGCACTTTTAGAAAAATAG 921  
 Db 907 CGTTTATTTTAGATAGAGAG 927

RESULT 12

AA330880

ID AAX30880 standard; DNA; 636 BP.

XX AAX30880;

XX 20-MAY-1999 (first entry)

Streptococcus pneumoniae genomic DNA sequence SEQ ID NO:157.

Streptococcus pneumoniae strain 0100993; vaccine; immune response;  
Streptococcal infection; pneumococcal; ss.

Streptococcus pneumoniae.

W09737026-A1.

09-OCT-1997.

01-APR-1997; 97WO-US053306.

22-AUG-1996; 96US-0025788.

02-APR-1996; 96US-0014690.

(SMIK) SMITHKLINE BEECHAM CORP.  
(SMIK) SMITHKLINE BEECHAM PLC.Black MT, Hodgson JE, Knowles DJC, Nicholas RO;  
Stodola RK;

WPI; 1997-503111/46.

P-PSDB; AAY11297.

Nucleic acids encoding pneumococcal polypeptide(s) - useful in  
vaccines, drug screening, etc

Claim 5; Page 150; 354pp; English.

AA330724 to AA330946 represent genomic DNA sequences isolated from  
Streptococcus pneumoniae strain 0100993. These genomic DNA sequences  
encode the novel proteins given in AAY11114 to AAY11367. The proteins,  
isolated from Streptococcus pneumoniae, can be used in vaccines against  
streptococcal infections and in assays for identifying compounds that

CC inhibit or activate the activity of the proteins. The antagonists can  
 CC be used to treat an individual having need to inhibit a bacterial  
 CC protein. Vectors expressing the proteins can be used to induce a  
 CC protective immune response in mammals.  
 XX  
 SQ Sequence 636 BP; 168 A; 144 C; 156 G; 168 T; 0 other;  
 Query Match 43.6%; Score 402; DB 18; Length 636;  
 Best Local Similarity 98.8%; Pred. No. 8.6e-117; Indels 0; Gaps 0;  
 Matches 405; Conservative 0; Mismatches 5;  
 QY 1 ATGACTAAACAGCCCTTTTATTTGCTGCTCAAGGTGCCAGTATCTAGGATGGGACGG 60  
 DB 85 ATGACTAAACAGCCCTTTTATTTGCTGCTCAAGGTGCCAGTATCTAGGATGGGACGG 144  
 QY 61 GATTTCATGATCAGTATCCGATGTCGAAGAACGATTCGAGCGAGTCAGGTGCTC 120  
 DB 145 GATTTCATGATCAGTATCCGATGTCGAAGAACGATTCGAGCGAGTCAGGTGCTC 204  
 QY 121 GGTATGATTTACGTTATCTCATGATACGGAAGAACAACTCAATCAGACCCGCTAT 180  
 DB 205 GGTATGATTTACGTTATCTCATGATACGGAAGAACAACTCAATCAGACCCGCTAT 264  
 QY 181 ACGCAACAGCCATCTAGGACCTCGGTTGCTATCTACCGTTTATTCGAAGAAAGGGC 240  
 DB 265 ACGCAACAGCCATCTAGGACCTCGGTTGCTATCTACCGTTTATTCGAAGAAAGGGC 324  
 QY 241 TATCAGCCTGATATGGTTGCTGTTGCTCTTGGAGAACTCTGCTTGGTGGCAGC 300  
 DB 325 TATCAGCCTGATATGGTTGCTGTTGCTCTTGGAGAACTCTGCTTGGTGGCAGC 384  
 QY 301 GCGCCTTGGATTTGAAGATCGGTTGCTTGGTAGCTAAGCGTGGAGCCTATATGGA 360  
 DB 385 GCGCCTTGGATTTGAAGATCGGTTGCTTGGTAGCTAAGCGTGGAGCCTATATGGA 444  
 QY 361 GAAGCGGCTCTGCTGACTCTGGCAGATGCTAGCAGTCTCTCATACGCC 410  
 DB 445 GAAGCGGCTCTGCTGACTCTGGCAGATGCTAGCAGTCTCTCATACGCC 494  
 RESULT 13  
 ABA90521  
 ID ABA90521 standard; DNA; 2365589 BP.  
 XX  
 AC ABA90521;  
 XX  
 DT 16-MAY-2002 (first entry)  
 XX  
 DE Genomic sequence of *Lactococcus lactis* IL1403.  
 XX  
 KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese; ds.  
 XX  
 OS *Lactococcus lactis* IL1403.  
 XX  
 PN FR2807446-A1.  
 XX  
 PD 12-OCT-2001.  
 XX  
 PF 11-APR-2000; 2000FR-0004630.  
 XX  
 PR 11-APR-2000; 2000FR-0004630.  
 XX  
 XX (INRG) INRA INST NAT RECH AGRONOMIQUE.  
 XX  
 PI Bolotine A, Sorokine A, Renault P, Ehrlich SD;  
 XX  
 DR WPI; 2002-043418/06.  
 XX  
 PT New nucleotide sequence useful in the identification of *Lactococcus*  
 PT *lactis* and related species -  
 XX  
 XX Claim 1; SEQ ID 1; 2504pp; French.  
 PS  
 XX

CC The present invention is related to a *Lactococcus lactis* nucleotide  
 CC sequence (ABA90521) and related proteins (ABA5300-ABA55621). The  
 CC nucleic acid sequence is useful in the detection and/or amplification of  
 CC nucleic acid sequence, particularly to identify *Lactococcus lactis* or  
 CC related species. The proteins of the invention are useful for the  
 CC biosynthesis or biodegradation of a composition of interest. The  
 CC invention helps research in lactic bacteria, particularly useful in the  
 CC production of yogurt and cheese.  
 CC Note: The sequence data for this patent is based on equivalent patent  
 CC WO2001/77334 (published 18-OCT-2001) which is available in electronic  
 CC format directly from WIPO at [ftp.wipo.int/pub/published\\_pat\\_sequences](http://ftp.wipo.int/pub/published_pat_sequences).  
 XX  
 SQ Sequence 2365589 BP; 765914 A; 415261 C; 420487 G; 763927 T; 0 other;  
 Query Match 42.2%; Score 388.8; DB 24; Length 2365589;  
 Best Local Similarity 64.6%; Pred. No. 1.2e-110; Indels 6; Gaps 1;  
 Matches 598; Conservative 0; Mismatches 322;  
 QY 1 ATGACTAAACAGCCCTTTTATTTGCTGCTCAAGGTGCCAGTATCTAGGATGGGACGG 60  
 DB 782872 ATGACTAAACAGCCCTTTTATTTGCTGCTCAAGGTGCCAGTATCTAGGATGGGACGG 782931  
 QY 61 GATTTCATGATCAGTATCCGATGTCGAAGAACGATTCGAGCGAGTCAGGTGCTC 120  
 DB 782932 GATTTCATGATCAGTATCCGATGTCGAAGAACGATTCGAGCGAGTCAGGTGCTC 782991  
 QY 121 GGTATGATTTACGTTATCTCATGATACGGAAGAACAACTCAATCAGACCCGCTAT 180  
 DB 782992 GGTATGATTTACGTTATCTCATGATACGGAAGAACAACTCAATCAGACCCGCTAT 783051  
 QY 181 ACGCAACAGCCATCTAGGACCTCGGTTGCTATCTACCGTTTATTCGAAGAAAGGGC 240  
 DB 783052 ACTCAACCTGCAATTTTAAACACCTCTGTTGCTATTTACGTTTGAAGTGAAGTGG 783111  
 QY 241 TATCAGCCTGATATGGTTGCTGTTGCTCTTGGAGAACTCTGCTTGGTGGCAGC 300  
 DB 783112 ATTAAACCTGACCTTGTGCTGCTCTTGTGTAATATCTGCTTGGTASCATCA 783171  
 QY 301 GCGCCTTGGATTTGAAGATCGGTTGCTTGGTAGCTAAGCGTGGAGCCTATATGGA 360  
 DB 783172 GGAATCATTTGATTTCAAGAACGATTAAGCTTGTAGCTAAACGCTGCTCAATATATGCA 783231  
 QY 361 GAAGCGGCTCTGCTGACTCTGGCAGATGCTAGCAGTCTCAATACGCCAGTACGAGTC 420  
 DB 783232 GAAGTGCACCGACTGGTTCTGGTAAATGGTGGCTGTAAATGAACACAGACCCAGCCTTG 783291  
 QY 421 ATTGAAGAACGCTGTCAAAAGCCTTCTGAACTT-----GGAGTGGTACTCCAGCCAC 474  
 DB 783292 ATTGAAGAACCTTGCAGAACGCGCGCAATTTAAAGGCGGTATTGTGAGTCCAGCAAT 783351  
 QY 475 TATAACACACCTGCAAAATCGTCAATGCTGGAGAGTGGTTGCAAGTTCAGTTCAGCGGTT 534  
 DB 783352 TATAACACACCGCCGCAAAATTTGTTGTTGTTGAGTGGAGCGGTTGATTTATGCTGTT 783411  
 QY 535 GAACCTTTTCAAGAACGAGTGTCCAAACGCTTGAATTCCTTAAAGTGTGAGTCCCTTTT 594  
 DB 783412 GAGTTGCTTAAAGAACGCGAGTTCGTAACCTATTGAATTTAAAGTTTCAGACCTTTC 783471  
 QY 595 CACACCGCTCTCTTGGAGCCTCTAGCCAGAACCTAGCTGAAACTCTAGCTCAGGTAACT 654  
 DB 783472 CATACAGCAATTTTAAACACGACATCTGAAATTTGGCTTTTGGAGCTTGATATAATGAT 783531  
 QY 655 TTTTCAGATTTTACTTGTCCCTAGTCGGCAATACAGAGCTGCTGTGATGCAAAAAGAG 714  
 DB 783532 TTTTAGACCTTTTGAATTACCATTAATCTCAATACAGTGTCTAAAGTANTGAAATGAT 783591  
 QY 715 GACATTGCTCAGCTCTTGAAGCGCTCAGGTCAAGAACCCGTTCTGTTCTATCAAGATATT 774  
 DB 783592 GAAGTCAAAGGACTTTTGACGCGTCAAGTCAAGTGAACAGTTCGTTTTTATCAATCGGTT 783651  
 QY 775 GGGGTATGCAAGACGACGATACCACTTTATCGAGATTGGACCGGGAAAGTCTTG 834  
 DB 783652 GAAACAAATGCAAAAACCTAGGGCGACTCGCTTTTATTGAAGTTGGTCTCTGGAGAGTACTT 783711

QY 835 TCAGGTTTCTTAAAAAATGATCAAACTGCTCACTAGCTCATGTGAAGATCAAGCG 894  
 DB 783712 TCAGGTTTCTTAAAAAATGATCAAACTGCTCACTAGCTCATGTGAAGATCAAGCG 894  
 QY 895 AGTTTAGTACACTTTTAGAAAAATA 920  
 DB 783772 TCATTGAAGCTTTGATTATCAGTA 783797

RESULT 14  
 ABN68640  
 ID ABN68640 standard; DNA; 924 BP.

XX AC ABN68640;  
 XX DT 01-JUL-2002 (first entry)  
 XX DE Streptococcus polynucleotide SEQ ID NO 5193.  
 XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 XX KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;  
 XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.  
 XX OS Streptococcus agalactiae.

XX WO200234771-A2.  
 XX PN 02-MAY-2002.

XX PF 29-OCT-2001; 2001WO-GB04789.  
 XX PR 27-OCT-2000; 2000GB-0026333.  
 XX PR 24-NOV-2000; 2000GB-0028727.  
 XX PR 07-MAR-2001; 2001GB-0005640.

XX PA (CHIR-) CHIRON SPA.  
 XX PA (GENO-) INST GENOMIC RES.

XX Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;  
 XX Tettelin H;  
 XX WPI; 2002-352536/38.  
 XX DR P-PSDB; ABP28009.

XX PT New Streptococcus protein for the treatment or prevention of infection  
 XX PT or disease caused by Streptococcus bacteria, such as meningitis, and  
 XX PT for detecting a compound that binds to the protein -  
 XX PS Claim 7; Page 3863; 4525pp; English.

XX CC The invention relates to a protein (ABP25413-ABP30895) from group B  
 XX CC streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS  
 XX CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 XX CC the specification. The proteins have antibacterial and antiinflammatory  
 XX CC activity. (I), nucleic acids encoding (I), ABN68644-ABN71526 and  
 XX CC antibodies that bind (I) are used in the manufacture of medicaments for  
 XX CC the treatment or prevention of infection or disease caused by  
 XX CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 XX CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
 XX CC biological sample. (I) is used to determine whether a compound binds to  
 XX CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
 XX CC used as a vaccine or diagnostic composition. The disease caused by  
 XX CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 XX CC acid encoding (I) may be used to recombinantly produce (I) and may be  
 XX CC used in gene therapy. Antibodies to (I) are used for affinity  
 XX CC chromatography, immunoassays, and distinguishing/identifying  
 XX CC Streptococcus proteins.

SQ Sequence 924 BP; 311 A; 132 C; 194 G; 287 T; 0 other;  
 Query Match 42.0%; Score 387; DB 24; Length 924;  
 Best Local Similarity 64.5%; Pred. No. 6e-112;

Matches 595; Conservative 0; Mismatches 325; Indels 3; Gaps 1;  
 QY 1 ATGACTAAAACAGAGCTTTTATTTGCTGTCAGAGTCCAGAGTATCTAGGATGGGACGG 60  
 DB 1 ATGAATAAGTTTCAATTTCTATTTGCTGGTCAAGGTGCTCAAAAAGTTAGGAATGGCTAGA 60  
 QY 61 GATTTCTATGATCAGTATCGATGTCAAAAGAAACGATTGATCGAGCGAGTCAGGTGCTC 120  
 DB 61 GATTTATATGAACCTTTCCCGATTTGTTAAAGAAACCTTTTATAGGCAAGTCATGTTTTA 120  
 QY 121 GGTATGATTTACGTTTATCTCATCGATACGGAAGAAACAACTCAATCAGACCGCGTAT 180  
 DB 121 GGATATGATTGAGAGAAATGATTGATAAGAGCTTAGACAAATTAATCAGAGAAATAT 180  
 QY 181 ACGCAACAGCCATCTAGAGCACTTCGGTTGCTATCTACCGTTTATTGCG---AAGAAAAG 237  
 DB 181 ACGCAACAGCTATTTCTAACGACTTCGACTGCTATTTATCGTTTAAATTTCTTAAAGAAAT 240  
 QY 238 GGCTATCAGCGCTGATATGTTGCTGTTGCTGTTGGAGAACTCTGCGCTTGGTGGCA 297  
 DB 241 GAACCTAGACAGATATGTTGAGGACTTTCTTTGGAGAAATACAGTGTCTTAGTTGCT 300  
 QY 298 AGCGGCGCTTGGATTGTTGAAGATGCGGTTGCTTGGTAGCTAAGCGTGGAGCCTATG 357  
 DB 301 TCAGGTGCTATAAGATTGGAAGATGCTGTAGTACTTCTGCTAGACGAGGCCAACTAATG 360  
 QY 358 GAAGAAGCGCTCCTGCTGCTGCTGCTGCAAGATGCTAGTCTCAATACGCCAGTAGAG 417  
 DB 361 GAACGAGCAGCACTGCTGTTAGTGGCAAAATGTTAGTACAGTTTAAATGCTGATCGTCAA 420  
 QY 418 GTCAATTGAAGAAGCTCTCAAAAAGCTTCTGAACCTTGGAGTGTACTCCAGGCAACTAT 477  
 DB 421 ATAATTGAAGATGCTTGTGAAGAGGCTTCTCAGTTTGGTATTGTTAGTCTCGCAATAT 480  
 QY 478 AACACACTCCACAATCGTCAITGCTGGAGAGTGGTTCAGTGTGATCGAGCGGTGAA 537  
 DB 481 AATACTCCAAAGCAAAATGTTTATAGGGGAGAGTCAATCGTCTGTAATGAGCTGTAGAA 540  
 QY 538 CTTTTTGAAGAAGCAGGTGCCAAACGCTTGAATTCCTCTTAAGGTGTGAGTCCCTTTTAC 597  
 DB 541 GAATTAACAACAAGAGGATAAACGTTTAATTCCTCTGAATGTTTCAGGTCTCTTTTAC 600  
 QY 598 ACCGCTCTCTTGAAGCTGTAGCCAGAACTAGCTGAAACTCTAGCTCAGTTAAGTTTT 657  
 DB 601 ACTGCTTTATGAAACCTGCAAGTCAAGAAATTGAGCGATGTTTATAGATAAAGTTCAATTT 660  
 QY 658 TCAGATTTTACTTGTCCCTAGTCCGCAATACAGAGCTGCTGTGATGCAAAAAGAGGAC 717  
 DB 661 TCAGTTTCTGAAATCCCTGTAAATTGGAAATACAGAGGCGCAGATAATGAAAAAGCGAT 720  
 QY 718 ATTGCTCAGCTCTTGACGCGTCAAGTCAAGAACCCGCTTCTGTTCTATGAAAGTATTGGG 777  
 DB 721 ATCAAAATCTTTATTTGGCACGCAAGTTATGGAACCTGTAGCTTTGATGAGTCTATTGAA 780  
 QY 778 GTCAATGAAGAAGCAGCAGCATAAGCACTTTATCGAGATTGGACCGGGGAAAGTCTTCTCA 837  
 DB 781 ACCATGAAAAATAATGGGAATGACACAGGTTGTTGAATTTGTTCCAGGSAAGTACTCTCA 840  
 QY 838 GGTTTGTTAAAAAATTTGATCAAACTGCTCACTTAGCTCATGTGGAGATCAAGCGAGT 897  
 DB 841 GGTTCCTTGAATAAATAGATAGTCTCATTCAGTACATTTCTGTGTAAGATAAAATTTGA 900  
 QY 898 TTAGTAGCACTTTTAGAAAAATA 920  
 DB 901 TTTAATAATCTTAAGGAATAAA 923

RESULT 15  
 ABN71527  
 ID ABN71527 standard; DNA; 2155561 BP.  
 XX AC ABN71527;  
 XX XX

DT 02-JUL-2002 (first entry)  
XX Streptococcus polynucleotide SEQ ID NO 10967.  
DE  
XX Streptococcus; GAS; group B streptococcus; Streptococcus agalactiae;  
KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;  
KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.  
XX Streptococcus sp.  
OS  
XX WO200234771-A2.  
PN  
XX 02-MAY-2002.  
PD  
XX 29-OCT-2001; 2001WO-GB04789.  
PF  
XX 27-OCT-2000; 2000GB-0026333.  
PR  
XX 24-NOV-2000; 2000GB-0028727.  
PR  
XX 07-MAR-2001; 2001GB-0005640.  
XX  
XX (CHIR-) CHIRON SPA.  
PA  
XX (GENO-) INST GENOMIC RES.  
PA  
XX Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;  
PI Tettelin H;  
PI  
XX WPI; 2002-352536/38.  
DR  
XX  
XX New Streptococcus protein for the treatment or prevention of infection  
PT or disease caused by Streptococcus bacteria, such as meningitis, and  
PT for detecting a compound that binds to the protein -  
PT  
XX  
XX Claim 8; Page 4196-4488; 4525pp; English.  
PS  
XX  
XX The invention relates to a protein (ABP25413-ABP30895) from group B  
CC Streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS  
CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in  
CC the specification. The proteins have antibacterial and antiinflammatory  
CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and  
CC antibodies that bind (I) are used in the manufacture of medicaments for  
CC the treatment or prevention of infection or disease caused by  
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
CC biological sample. (I) is used to determine whether a compound binds to  
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
CC used as a vaccine or diagnostic composition. The disease caused by  
CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
CC acid encoding (I) may be used to recombinantly produce (I) and may be  
CC used in gene therapy. Antibodies to (I) are used for affinity  
CC chromatography, immunoassays, and distinguishing/identifying  
CC Streptococcus proteins.  
XX  
SQ Sequence 2155561 BP; 695741 A; 385320 C; 382958 G; 691542 T; 0 other;  
Query Match 42.0%; Score 387; DB 24; Length 2155561;  
Best Local Similarity 64.5%; Pred. No. 4.3e-110;  
Matches 595; Conservative 0; Mismatches 325; Indels 3; Gaps 1;  
QY 1 ATGACTAAACAGCGCTTTTATTTGCTGGTCAAGTGCCCGAGTATCTAGGGATGGGACGG 60  
DB ATGAATAAAGTTTCATTTCTATTTGCTGGTCAAGTGCTCAAAAGTTAGGAATGGCTAGA 331006  
QY 61 GATTTCATGATCAGTATCCGATTTGCTCAAGAGACGATTCGATCAGCGAGTCAGTCTC 120  
DB GATTATAGAACTTTCCCGATTTGTTAAAGAACTTTTATGATAGGCAAGTCATGTTTTA 331066  
QY 121 GGTATGATTTACGTTATCTCATCGATACGGAAGACAAACTCAATCAGACCCGCTAT 180  
DB GGATATGATTTGAGAGAATTCGATTTGATTAAGACCTTAGCAAAATTAATCAGACGAATAT 331126  
QY 181 ACGCAACAGCATTCTAGGACATCGGTGCTATCTACCGTTTATTCG---AGAAAAAG 237  
DB ACGCAACAGCTATTTCTAAACGACTTCGACTGCTATTTATCGTTTAATTTCTTAAGAAAT 331186

QY 238 GGCTATCAGCCTGATATGTTGCTGTTGTTCTCTTGGAGAACTACTCTCCTTGGTGGCA 297  
DB 331187 GAACTTAGACCAGATATGTTGCTGAGACTTTCTTTGGGAGAAATACAGTGTCTTTAGTTGCT 331246  
QY 298 AGCGGCGCTTGGATTTTGAAGATCGGTGCTTGGTGTAGCTTAAGCGTGAGCCCTATATG 357  
DB 331247 TCAAGTGTCTATAGATTTGAAGATCTGTAGTACTTTGCTGTAGACGAGGCCCACTAAATG 331306  
QY 358 GAAGAACGCGCTCCTGCTGACTCTGGCAAGATGGTAGCAGTTCTCAATACGCCAGTAGAG 417  
DB 331307 GAAGCAGCAGCACCTGCTGCTAGTGGCAAAATGGTAGCAGTTTAAATCTCATCGTCAA 331366  
QY 418 GTCAITGAAGAAGCCCTGTCAAAAGCTTCTGAACCTTGGAGTGGTTTACTCCAGCCACATAT 477  
DB 331367 ATAATTGAAGATGCTTTGTAAGAAGGCTTCTCAGTTTGGTATTGTTAGTCTCTCAAAATTA 331426  
QY 478 AACACACCTTGCACAAATCGTCAATGCTGGAGAAGTGGTTGACAGTTGATCGAGCGGTGAA 537  
DB 331427 AATACTCCAAAGCAAAATGTTATAGGGGAGAGTCAATCGCTGTAAATCGACGTGTAGAA 331486  
QY 538 CTTTTCGAAGAAGCAGGTGCCAAACGCTTGATTCCTCTTAAGGTGTGAGGTCCCTTTTCAC 597  
DB 331487 GAATTAACCAACAAGGAGTAAACGTTTAAATCTCTGAATGTTTTCAGGTCCCTTTTCAC 331546  
QY 598 ACCGCTCTCCTTGAGCCTGCTAGCCAGAAACTAGCTGAAACTCTAGCTCAGGTAAAGTTT 657  
DB 331547 ACTGCTTTATTGAACCTGCAAGTCAAGATTTAGCGGATGTTTATAGATAAAGTTTCA 331606  
QY 658 TCAGATTTTACTTTGCTCCCTAGTTCGGCAATACAGAGCTGTGTGATGCAAAAAGAGGAC 717  
DB 331607 TCAGTTTCTGAAATCCCTGTAATTTGGAATACAGAGGCGGCAGATAATGAAAAAGACCAT 331666  
QY 718 ATTGCTCAGCTCTTGACGGGTGCTCAGGTCAAGGAACCGTTTGGTTTCTATGAAAGTATTGG 777  
DB 331667 ATCAAACTTTTATTTGGCAGCCCAAGTTATGGAACCTGTACGCTTTGATGAGTCTATTGAA 331726  
QY 778 GTCATGCAAGAAGCAGGCGATTAAGCAACTTTATCGAGATTGGACCGGGGAAAGCTTTGTCA 837  
DB 331727 ACCATGAAAAAATGGAATGACACAGGTTGTTGAAATTTGGTCCAGGGAAGGTACTCTCA 331786  
QY 838 GGTTTTGTAAAAAATTTGATCAAACTGCTACCTTAGCTCATGTGGAAGATCAAGGAGT 897  
DB 331787 GGTTCCTTGAAGAAAAATAGATAGCTCATTATCAGTACATCTCTGTTGAAGATAAAAATTGGA 331846  
QY 898 TTAGTAGCAGCTTTTAGAAAAATA 920  
DB 331847 TTTAATAATCTTAAGGAATTTAA 331869

Search completed: June 11, 2003, 18:04:08  
Job time : 275 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 11, 2003, 17:59:32 ; Search time 70 Seconds  
(without alignments)  
4034.990 Million cell updates/sec

Title: US-09-308-397-1  
Perfect score: 921  
Sequence: 1 atcactaaaacagccttttt.....tagcacttttagaaaaaatag 921

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTUS COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	913	99.1	19702	4	US-08-961-527-7
2	523.2	56.8	1196	4	US-08-858-207A-8
C 3	164.8	17.9	9686	4	US-09-221-017B-1003
4	130.4	14.2	936	4	US-09-134-001C-1987
5	126.4	13.7	927	1	US-08-789-609A-1
6	126.4	13.7	927	4	US-09-108-517-1
7	69.2	7.5	19227	3	US-09-090-793-13
8	69.2	7.5	40138	3	US-09-090-793-12
9	53.4	5.8	8268	1	US-08-375-709-10
10	53.4	5.8	8268	1	US-08-752-929-10
11	53.4	5.8	37895	1	US-08-375-709-1
12	53.4	5.8	37895	1	US-08-752-929-1
13	53.4	5.8	37895	3	US-09-090-793-1
14	53.2	5.8	4403765	4	US-09-103-840A-2
15	53.2	5.8	4411529	4	US-09-103-840A-1
16	47.2	5.1	80161	3	US-09-036-987A-1
17	47.2	5.1	80161	4	US-09-370-700-1
18	45	4.9	345	3	US-09-154-083-9
C 19	40.4	4.4	4403765	4	US-09-103-840A-2
C 20	40.4	4.4	4411529	4	US-09-103-840A-1
C 21	40.2	4.4	2190	4	US-09-625-188-19
22	38.6	4.2	925	3	US-08-858-003-1
23	38.6	4.2	925	3	US-09-078-166-1
24	38.6	4.2	925	4	US-08-997-467-1
25	38.2	4.1	1245	3	US-08-242-050-1
26	38.2	4.1	19702	4	US-08-961-527-7
27	36.6	4.0	15872	4	US-09-105-537-1

28 36.6 4.0 50937 4 US-09-428-517-1 Sequence 1, Appli  
29 35.8 3.9 71989 4 US-09-443-501A-2 Sequence 2, Appli  
30 35.6 3.9 3279 5 PCT-US93-03077-2 Sequence 2, Appli  
31 35.4 3.8 13987 2 US-08-804-227C-13 Sequence 13, Appli  
32 35.4 3.8 44377 2 US-08-804-227C-7 Sequence 7, Appli  
33 35.4 3.8 44377 2 US-08-804-198-1 Sequence 1, Appli  
34 35.2 3.8 13842 4 US-09-105-537-30 Sequence 30, Appli  
35 35.2 3.8 36778 4 US-09-105-537-5 Sequence 5, Appli  
36 35.2 3.8 38506 3 US-09-320-878-19 Sequence 19, Appli  
C 37 34.8 3.8 2824 4 US-07-757-022B-13 Sequence 13, Appli  
C 38 34.8 3.8 3066 4 US-07-757-022B-83 Sequence 83, Appli  
C 39 34.8 3.8 3117 4 US-07-757-022B-73 Sequence 73, Appli  
C 40 34.8 3.8 3148 4 US-07-757-022B-57 Sequence 57, Appli  
C 41 34.8 3.8 3420 4 US-07-757-022B-103 Sequence 103, Appli  
C 42 34.8 3.8 3813 4 US-07-757-022B-43 Sequence 43, Appli  
C 43 34.8 3.8 3936 4 US-07-757-022B-41 Sequence 41, Appli  
C 44 34.8 3.8 3942 4 US-07-757-022B-141 Sequence 141, Appli  
C 45 34.8 3.8 3945 4 US-07-757-022B-49 Sequence 49, Appli

## ALIGNMENTS

## RESULT 1

US-08-961-527-7/c

; Sequence 7, Application US/08961527

; Patent No. 6420135

; GENERAL INFORMATION:

; APPLICANT: Charles Kunsch

; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

; NUMBER OF SEQUENCES: 391

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/961.527

; FILING DATE:

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Brookes, A. Anders

; REGISTRATION NUMBER: 36,373

; REFERENCE/DOCKET NUMBER: PB340P1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 309-8504

; TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 19702 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

US-08-961-527-7

Query Match 99.1%; Score 913; DB 4; Length 19702;

Best Local Similarity 99.5%; Pred. No. 8.1e-295;

Matches 916; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ATGACTAAACAGCCCTTTTATTTCTGCTCAAGTGCCCGAGTATCTAGGATGGGACGG 60

Db 17149 ATGACTAAACAGCCCTTTTATTTCTGCTCAAGTGCCCGAGTATCTAGGATGGGACGG 17090

```
QY 61 GATTCTATGATCAGTATCCGATTGTCAAAGAAAGATTGATCGAGCGAGTCAGGTGCTC 120
Db 17089 GATTCTATGATCAGTATCCGATTGTAAAGAAAGATTGATCGAGCGAGCGAGTCAGGTGCTC 17030
QY 121 GATTATGATTTAGTTATCTATCATCATGATACGGAAGAAAGAACTCAATCAGACCGCGTAT 180
Db 17029 GATTATGATTTAGTTATCTATCATCATGATACGGAAGAAAGAACTCAATCAGACCGCGTAT 16970
QY 181 AGCGACAGACCATCTAGCGACTTCGGTGTGATCTACCGTTTATTGCAAGAAAGAGGC 240
Db 16969 AGCGACAGACCATCTAGCGACTTCGGTGTGATCTACCGTTTATTGCAAGAAAGAGGC 16910
QY 241 TATCAGCCTGATATGTTGCTGTTGCTCTCTGAGAAATCTCTGCTTGGTGGCAAGC 300
Db 16909 TATCAGCCTGATATGTTGCTGTTGCTCTCTGAGAAATCTCTGCTTGGTGGCAAGT 16850
QY 301 GCGCGCTGGATTGAGATCGGCTGCTGTTGTTAGCTGAGCGTGGAGCGCTATATGAA 360
Db 16849 GCGCGCTGGATTGAGATCGGCTGCTGTTGTTAGCTGAGCGTGGAGCGCTATATGAA 16790
QY 361 GAAGCGGCTCCTGCTGACTCTGCGCAAGATGTTAGCAGTTCTCAATACGCCAGTAGAGTC 420
Db 16789 GAAGCGGCTCCTGCTGACTCTGCGCAAGATGTTAGCAGTTCTCAATACGCCAGTAGAGTC 16730
QY 421 ATTGAAGAAGCCTGTCAAAAAGCTTCTGNACTTGAGTGGTTACTCCAGCGCAACTATAAC 480
Db 16729 ATTGAAGAAGCCTGTCAAAAAGCTTCTGNACTTGAGTGGTTACTCCAGCGCAACTATAAC 16670
QY 481 ACACCTGCAACAATGCTCAATGCTGGAGAAAGTGGTTGCAAGTTGATCGAGCGGTTGAACTT 540
Db 16669 ACACCTGCAACAATGCTCAATGCTGGAGAAAGTGGTTGCAAGTTGATCGAGCGGTTGAACTT 16610
QY 541 TTGCAAGAAGCAGGTGCCAAACGCTTGATTCCTTAAGGTGTAGGTCCCTTTACACACC 600
Db 16609 TTGCAAGAAGCAGGTGCCAAACGCTTGATTCCTTAAGGTGTAGGTCCCTTTACACACC 16550
QY 601 GCTCTCTTTGAGCGCTCTAGCAGAACTAGCTGAACTCTAGCTCAGGTAAAGTTTTC 660
Db 16549 GCTCTCTTTGAGCGCTCTAGCAGAACTAGCTGAACTCTAGCTCAGGTAAAGTTTTC 16490
QY 661 GATTATCTGTCCTCTAGTCTGCGCAATACAGAACTGCTGATGCAAAAAGAGACATTT 720
Db 16489 GATTATCTGTCCTCTAGTCTGCGCAATACAGAACTGCTGATGCAAAAAGAGACATTT 16430
QY 721 GCTCAGCTCTGAGCGGTGAGGTCAAGGAACCGTTCTGTTCTATGAAGTATTCGGGTC 780
Db 16429 GCTCAGCTCTGAGCGGTGAGGTCAAGGAACCGTTCTGTTCTATGAAGTATTCGGGTC 16370
QY 781 ATGCAAGAAGCAGGCAATAAGCACTTATCGAGATGGACCGGGGAAAGTCTTTGTCAGGT 840
Db 16369 ATGCAAGAAGCAGGCAATAAGCACTTATCGAGATGGACCGGGGAAAGTCTTTGTCAGGT 16310
QY 841 TTTGTTAAAAAATTGATCAAACTGCTCACTAGCTATGTCGAGATCAAGCGAGTTTA 900
Db 16309 TTTGTTAAAAAATTGATCAAACTGCTCACTAGCTATGTCGAGATCAAGCGAGTTTA 16250
QY 901 GTAGCACTTTTGAAGAAATAG 921
Db 16249 GTAGCACTTTTGAAGAAATAG 16229
```

## RESULT 2

```
US-08-858-207A-8
; Sequence 8, Application US/08858207A
; Patent No. 6348328
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Richard
; APPLICANT: Stodola, Robert
; TITLE OF INVENTION: No. 6348328el Compounds
; NUMBER OF SEQUENCES: 552
```

```
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,207A
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/017670
; FILING DATE: 14-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R
; REGISTRATION NUMBER: 36,891
; REFERENCE/DOCKET NUMBER: P50475
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1196 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-858-207A-8
```

```
Query Match 56.8%; Score 523.2; DB 4; Length 1196;
Best Local Similarity 99.1%; Pred. No. 4.9e-165;
Matches 547; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 372 TGCTGACTCTGCGCAAGATGTTAGTACGTTCTCAATAGCCAGTAGAGGTTCATTGAAAGC 431
Db 1 TGCTGAACTTTGGCAAGATGTTAGTACGTTCTCAATAGCCAGTAGAGGTTCATTGAAAGC 60
QY 432 CTGTCAAAAAGCTTCT-GAACTTTGGAGTGGTTACTCCAGCCAACTATAACACACTGCAC 490
Db 61 CTGTCAAAAAGCTTCTGGAACCTTGGAGTGGTTACTCCAGCCAACTATAACACACTGCAC 120
QY 491 AAATCGTCATGCTGGAGAAAGTGGTTGAGTTCATCGAGCGGTGAACTTTGCAAGAAG 550
Db 121 AAATCGTCATGCTGGAGAAAGTGGTTGAGTTCATCGAGCGGTGAACTTTGCAAGAAG 180
QY 551 CAGGTGCAAAACGCTTCTTAAAGTGTGAGTCCCTTTCAACCGCTCTCCTTG 610
Db 181 CAGGTGCAAAACGCTTCTTAAAGTGTGAGTCCCTTTCAACCGCTCTCCTTG 240
QY 611 AGCTGTAGCCAGAACTAGCTGAAACTAGCTCAGGTAAAGTTCAGATTTTACTT 670
Db 241 AGCTGTAGCCAGAACTAGCTGAAACTAGCTCAGGTAAAGTTCAGATTTTACTT 300
QY 671 GTCCCTCTAGTCGGCAATACAGAAAGTCTGTGATGCAAAAAGAGGACATTTGCTCAGCTCT 730
Db 301 GTCCCTCTAGTCGGCAATACAGAAAGTCTGTGATGCAAAAAGAGGACATTTGCTCAGCTCT 360
QY 731 TGAGCGCTCAGGTCAAGAAACCGTTCGTTTCTTATGAAAGTATTTGGGTTCATCAAGAAG 790
Db 361 TGAGCGCTCAGGTCAAGAAACCGTTCGTTTCTTATGAAAGTATTTGGGTTCATCAAGAAG 420
QY 791 CAGGCATAAGCAACTTTA-TCGAGATTGGACCGGGGAAAGTCTTTGTGAGTTTGTATAA 849
Db 421 CAGGCATAAGCAACTTTA-TCGAGATTGGACCGGGGAAAGTCTTTGTGAGTTTGTATAA 480
QY 850 AAAATTGATCAAACTGCTCACTTAGCTCATGTGGAAGATCAACCGAGTTTAGTAGACTT 909
```

Db 481 AAAATTGATCAAACTGCTCACTTAGCTCATGTGGAAGATCAAGCGAGTTTAGTAGCACTT 540

QY 910 TTAGAAAATAG 921

Db 541 TTAGAAAATAG 552

RESULT 3

US-09-221-017B-1003/c

; Sequence 1003, Application US/09221017B

; Patent No. 6444799

; GENERAL INFORMATION:

; APPLICANT: Ross, Bruce C.

; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF

; NUMBER OF SEQUENCES: 1120

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORRISON & FOERSTER

; STREET: 755 PAGE MILL ROAD

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304-1018

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Windows

; SOFTWARE: FastSEQ for Windows Version 2.0b

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/221.017B

; FILING DATE: 23-DEC-1998

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PP1182

; FILING DATE: 31-DEC-1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PP1546

; FILING DATE: 30-JAN-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PP2911

; FILING DATE: 09-APR-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/AU98/01023

; FILING DATE: 10-DEC-1998

; ATTORNEY/AGENT INFORMATION:

; NAME: Monroy, Gladys H

; REGISTRATION NUMBER: 32,430

; REFERENCE/DOCKET NUMBER: 27340-20021.00

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-813-5600

; TELEFAX: 650-494-0792

; INFORMATION FOR SEQ ID NO: 1003:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 9686 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: circular

; MOLECULE TYPE: DNA (Genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: UNKNOWN

; ORIGINAL SOURCE:

; ORGANISM: PORPHYROMONAS GINGIVALIS

; NAME/KEY: misc feature

; LOCATION: 1...9686

US-09-221-017B-1003

Query Match 17.9%; Score 164.8; DB 4; Length 9686;

Best Local Similarity 52.8%; Pred. No. 2e-44;

Matches 452; Conservative 0; Mismatches 392; Indels 12; Gaps 4;

QY 10 ACAGCCTTTTATTGCTGCTCAGGTGCCAGTATCTAGGATGGACGGGATTTCAT 69

Db 4537 AAAGCATTTGTTATCCCGGTTCAGGAGCACAGTTCTGTAGGAATGGGAAAAGACCTGTAC 4478

QY 70 GATCAGTATCCGATTGTCAAGAAACGATTGATCAGCGAGTCAAGTCTCGTTATGAT 129

Db 4477 GAACAGAACCCCGAGGCCAAGCATATTTTGAAGAACCAATGAGATATTAGGCTTCGC 4418

QY 130 TTACGTTATCTCATCATGATACGGAA---GAAGACAACTCAATCAGACCCGCTATACGGAA 186

Db 4417 ATTACGGACATCCTTTCAACGGCACGGCAGAGGAGCTGAACAAACAACTGACCCAG 4358

QY 187 CCAGCCATTCTAGCACTTCGTTTCTATCTACCTGTTTATTGCAAGAAAAGGGCTATCAG 246

Db 4357 CGGCGATCTTCTCCACTCGTGATTCTGGCAAGACGATGGGCGAAGA---CTTCCT 4301

QY 247 CTTGATATGTTGCTGTTTGTCTCTTGSAGAACTCTGCTTGTGTCGCAAGCGGCC 306

Db 4300 CCCGATATGGTGGCAGGCCATTCGCTCGGAGAGTTTTCGGCATTTGTTGCGACAGGTGCC 4241

QY 307 TTGGATTTTGAAGATGCGGTTCCTTGGTAGCTAAGCGTGGAGCCTATATGAAAGACG 366

Db 4240 ATGACCTTCGAAGATGTTCTTCGTTCTGATCGAAACGTCCTGCCATGCAGAAAGCA 4181

QY 367 GCTCTGCTGACTCTGGCAAGATGATGAGTCTCTCAATACGCCAGTAGAGTCAATTGAA 426

Db 4180 TCGAAGTACGCCCTCGACTATGCGGCTGTACTGGGTCTGCTGATGAAAGGTGAA 4121

QY 427 GAAGCTGTCAAAAAGCTTCTGAACCTTGGAGTGGTACTCCAGCCAACTATAACACCT 486

Db 4120 GAAATCTGTGCCCAAGTACCGA---TGAAGTTGCTACCGGCCAACTACACTGCCCG 4064

QY 487 GCACAAATCGTCAATGCTGGAGAGTGGTTGAGTTGATCGAGCGGTTGAACTTTTGCAA 546

Db 4063 GGACAGATCGTTATATCCGGTTCGGTAGAAGAGTGGACAGAGCTTGGCAACTGCTCAAG 4004

QY 547 GAAGCAGTGGCAACGCTTGATTCCTCTTAAGTGTGAGTCCCTTTTACACCGCTCTC 606

Db 4003 GAAGCCGAGCCAGCGTGTCTTAAAGCTCGCTGCGTGAGCAATTCATTCCTCTG 3944

QY 607 CTTGAGCCTGTAGCCAGAAACTAGCTGAAACTTAGCTCAGGTAAAGTTTTTCAGATTT 666

Db 3943 ATGGAACCGGCTCGTGAAGAGCTGSCAAAGCCATCAGGAGACTACGATCAGTCAGCG 3884

QY 667 ACTTGTCCCTAGTCGGCAA---TACAGAACTGCTGTGATGCAAAAAGAGACATTGCT 723

Db 3883 ATCTGCCCATCTATCAGAAAGTAAAGCGCTTCTGCCGTCACCGATCCGCGCAGAGATTAA 3824

QY 724 CAGCTTTGACGCGTCAGTCAAGAAACCGTTCGTTTCTATGAAAGTATTGGGTCATG 783

Db 3823 AAGAACCTGTAGACACAGCTGACGCGACCGGTTCCCTGGACACAGAGTGTGCTGAATG 3764

QY 784 CAAGAAGCAGGATGAAGCAATTTATCGAGATTGACCGGGGAAAGTCTTGTGAGTTT 843

Db 3763 ACAGCCGATGTGCGCGACCACTTTATGGAGTTAGGCCCGGAAATGTATTGAGGGTTG 3704

QY 844 GTTAAAAAATGTATC 859

Db 3703 GTGAAGAAAAATGCTC 3688

RESULT 4

US-09-134-001C-1987

; Sequence 1987, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

```
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1987
; LENGTH: 936
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1987

Query Match      14.2%; Score 130.4; DB 4; Length 936;
Best Local Similarity 50.1%; Pred. No. 1.5e-33;
Matches 438; Conservative 0; Mismatches 421; Indels 15; Gaps 4;

QY 1 ATGACTAAACAGCCTTTTATTGCTGCTCAAGTGCCAGTACTAGGATGGGCGG 60
DB 10 AUGGGTAAATAGCAGTTGCTCTCCCTGGCCAAAGTGACAAAAAGTAGGATGGCTAGT 69
QY 61 GATTTCATGATCAGTATCCGATTTGTCAAGAAACGATTGATCGAGCGAGTCAGGTGCTC 120
DB 70 GACTTATATAATGAAGAGAACGCTCGACTCGAGTATTAAATTTAGCTCAAGAGACAGTA 129
QY 121 GGTATGATTAGGTATCTCATCATGATAGGAGAGAC---AAACTCAATCAGNCCGC 177
DB 130 GACTTTGATTGCTAGAAACGATGTTACTGATAATGATGGAAATTAGGAGAAACAGAA 189
QY 178 TATACCAACACGACCTTCTAGCGACTTCGGTGTGCTATCTACCGTTTATTGCAAGAAAAG 237
DB 190 AATACACACCTGCCTATTACACATAGTGTGCGTTACTAGATCATTAATCATTTA 249
QY 238 GGGTATCAGCCTGATATGTTGCTGCTGTTGCTCTTGGAGATATCTGCTGCTGGTGCA 297
DB 250 AAGTAGATATATCTATGCGACATAGCTTA-----GGAGATATTCAGCTTAGTTTCT 303
QY 298 AGCGGGCCCTGGATTTCGAAGATCGGTTGCTTGGTAGCTTAAGCTGGAGCCTATATG 357
DB 304 ACTGATGATTATCTTTGAAATGCGGTAGAAATGTCGTAACGTTGGCCAACTATG 363
QY 358 GAAGAAAGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 417
DB 364 GCTCAAGCGTTTCTTAACGCTGTGGAGGTATGGCAGCAGTATTAGTTTGGATTATGAT 423
QY 418 GTCAATTGAAGAGCCTGCTCAAAAGCTCTGAACTTGGAGTGGTTACT---CCAGCCAC 474
DB 424 GATGTTGATAGATATGCTCAACGTTATCTCAAAAGAACAGTTAATTAACCTGCTAAT 483
QY 475 TATAACACCTGCTCAAAATCGTCAATGCTGAGAGAGTGGTTGCAAGTTCATTCGAGCGGTT 534
DB 484 ATTAACCTACACAGGTCAAAATCGTGTGCTGCAATATAATCTTTAATTGATGAATTAGTA 543
QY 535 GACTTTTGAAGAGCAGGTGCTCAACGCTTGAATTCCTTGAAGTCTCAGTCCCTTT 594
DB 544 GAAAGGGCAAGAACCTTGGTCTAAACGCTGCTTCCATTAGCTGTTTCCGGTCCCTTT 603
QY 595 CACACCGCTCTCCTTGAGCCTGCTAGCCAGAACTAGCTGAAACTCTAGCTCAGGTAAGT 654
DB 604 CATCTCTCAATGATGAAGTATTGAAGAGGATTTTGCTAATTTCAATTAATCAATTGAA 663
QY 655 TTTTCAGATTTTACTTGCTCCCTAGTCGCAATACAGAGCTGCTGATGCAAAAGAG 714
DB 664 TGCATAATGCTAAATATATCCAGTTGTTCAAGATGTTAATGCAAGGGAGAAACCGATGCT 723
QY 715 GACATTGCTCAGCTCTGACGGT---CAGGTCAAGGAACCGTTGCTTCTTATGAAAGT 771
DB 724 GAAGTAATTAACGCAATATGTTAAACAAATTAATTCACCTGTTCAATTTATTCATCA 783
QY 772 ATGGGGTCAATCAAGAGCAGGCTAAGCAACTTTATCGAGATTGACCGGGGAAAGTC 831
DB 784 ACGAGTGGTTGATTAATCAAGGTGTGCAATCACTTTATTGAAATGGACCGGAAAGTA 843
QY 832 TTGTCAGGTTTGTAAAAAATTTGATCAAACTG 865
DB 844 TTAICTGGGCTTATCAAAAAAATAAATCGAGATG 877
```

```
RESULT 5
US-08-789-609A-1
; Sequence 1, Application US/08789609A
; Patent No. 5827689
; GENERAL INFORMATION:
; APPLICANT: Gentry, Daniel
; APPLICANT: Lonsdale, John
; APPLICANT: Pearson, Stewart
; APPLICANT: Payne, David
; TITLE OF INVENTION: No. 5827689el FabD
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/789,609A
; FILING DATE: 24-JAN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: 60/030685
; APPLICATION NUMBER: 60/030685
; FILING DATE: 13-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: GM50004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5050
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 927 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-08-789-609A-1
```

```
Query Match      13.7%; Score 126.4; DB 1; Length 927;
Best Local Similarity 49.1%; Pred. No. 3.2e-32;
Matches 455; Conservative 0; Mismatches 456; Indels 15; Gaps 4;

QY 1 ATGACTAAACAGCCTTTTATTGCTGCTCAAGTGCCCGAGTATCTAGGGATGGGACGG 60
DB 1 ATGAGTAAACAGCAATTTATTTCCGGGACAAAGGTGCCCAAAAGTTGGTATGGCACA 60
QY 61 GATTTCATGATCAGTATCGGATTCGAAGAAACGATTGATCGAGCGAGTCAGGTGCTC 120
DB 61 GATTGTTTAAATAACAATGATCAAGCAACTGAAATTTTAACCTCAGCAGCAAGACGTTA 120
QY 121 GGTATGATTACGTTATCTCATCGATAGC---GAAGAAAGACAAACTCAATCAGACCCGC 177
DB 121 GACTTTGATATTTTAGAGACAATGTTTACTGATGAAGAGGTAAATTTGGTGAACAGTAA 180
QY 178 TATACGCAACACGCAATCTTAGGACTTCGGTTGCTATCTACCGTTTATTGCAAGAAAG 237
DB 181 AACACGCAACCAAGCTTTATTGACGCATAGTTCGGCATTA-----TTAGCAGCGCTAAAA 234
QY 238 GGTATCAGCCTCATATGTTGCTGTTGTTGTTCTCTGGAGAAATCTCTGCCCTTGGTGCA 297
DB 235 ATTTTGAATCCTGATTTTACTATGCGGCATAGTTTAGGTGAATATCAAGTTAGTTGCA 294
QY 298 AGCGGCGCTTGGATTTTGAAGATCGGTTGCTTGGTAGCTTGGTGAAGCGTGAAGCTATATG 357
```

Db 295 GCTGACGTATTATCATTTTGAAGATGCGAGTTAAATTTGTAGAAAACGTTGTCATTAATG 354  
 Qy 358 GAAGAGCGGCTCTGCTGACTCTGGCAGATGCTAGCAGTTCTCAATAGCCAGTAGAG 417  
 Db 355 GCGCAAGCATTTCTACTGTTGTAGGAGCATGCTGCTGAGTATTTGGGATTTAGATTTT 414  
 Qy 418 GTCAATTGAAGAGCTGTCA---AAAAGCTTCTGAACTTTGGAGTGGTTACTCCAGGCAAC 474  
 Db 415 AAAGTCGATGAATTTGTAGTCATTTATCATCTGATGACAAATAATTTGAACGCAAC 474  
 Qy 475 TATAACACCTGACAAATCGTCAATTTCTGCTGGAGAGTGGTTGAGTGTATGAGCGGTT 534  
 Db 475 ATTAAATGCCAGTCAAAATTTGTTTCAAGTCCAAAGCTTTAAATTTGATGAGTAGTA 534  
 Qy 535 GAACTTTTGCAGAGCGTCCAAAGCGTTGATTTCTCTTAAGGTCAGGTCCCTTT 594  
 Db 535 GAAAAGGTAATCATTTAGTGCATTTGCTGATGACAAATAATTTGAACGCAAC 594  
 Qy 595 CACACCGCTCTCTTGGAGCGTGTAGCAGAACTAGTGAATCTAGTCTAGGTAAGT 654  
 Db 595 CATTCATCGCTAATGAAAGTGAATGAAGAGATTTTCAAGTTACATTAATCAATTGAA 654  
 Qy 655 TTTTCAGATTTTACTTCTCCCTAGTGGCAATACAGAGCT---GCTGTGATGCAAAA 711  
 Db 655 TGGCGTGAAGTAAAGTTCTCTGATTTCAAAATGTAATGCGCAAGTGAATGACAAA 714  
 Qy 712 GAGGACATTTGCTAGCTCTTGACGCGTCAGTCAAGGAAACCGTTTCTGTTCTATGAAAGT 771  
 Db 715 GAAGTAATTAATCTAATATGTTCAAGCAATATATTTACCAGTACATTAATCAATCA 774  
 Qy 772 ATTGGGTCATGCAAGAGCAGGATAGCAACTTTTACGAGATGACCGGGGAAAGTC 831  
 Db 775 ACAGAAATGGCTAATAGACCAAGTGTGATCAITTTTATTGAAATTTGCTCTGGAAGTT 834  
 Qy 832 TTGTGAGGTTTGTAAATAATGATCAAACTGCTCACTAGTCTCATGTGGAAGATCA 891  
 Db 835 TTATCTGCTAATTAATAATAATAGAGATTTAATGATTAATCAATTCAACTTTA 894  
 Qy 892 GCGAGTTAGTACACTTTTAGAAAA 917  
 Db 895 GAAGATGGAAGGATGGAATGAAA 920

RESULT 6

US-09-108-517-1  
 ; Sequence 1, Application US/09108517  
 ; Patent No. 6258934  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gentry, Daniel  
 ; APPLICANT: Lonsdale, John  
 ; APPLICANT: Pearson, Stewart  
 ; APPLICANT: Payne, David  
 ; TITLE OF INVENTION: No. 6258934el PabD  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: SmithKline Beecham Corporation  
 ; STREET: 709 Swedeland Road  
 ; CITY: King of Prussia  
 ; STATE: PA  
 ; COUNTRY: USA  
 ; ZIP: 19406-0939  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/108,517  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/789,609  
 ; FILING DATE:

ATTORNEY/AGENT INFORMATION:  
 NAME: Gimmli, Edward R  
 REGISTRATION NUMBER: 38,891  
 REFERENCE/DOCKET NUMBER: GMS0004  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 610-270-4478  
 TELEFAX: 610-270-5090  
 TELEX:  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 927 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: Genomic DNA  
 US-09-108-517-1

Query Match 13.7%; Score 126.4; DB 4; Length 927;  
 Best Local Similarity 49.1%; Pred. No. 3.2e-32;  
 Matches 455; Conservative 0; Mismatches 456; Indels 15; Gaps 4;  
 Qy 1 ATGACTAAAACAGCCCTTTTATTTCTGCTGCTCAAGTGGCCAGTATCTTAGGGATGGACGG 60  
 Db 1 ATGAGTAAAACAGCAATTTATTTTCCGGACAAAGTGCCAAAAGTTGGTATGGCAAA 60  
 Qy 61 GATTTCTATGATCAGTATCCGATTTCTCAAGAAACGATTGATCGAGCGAGTCAGGTGCTC 120  
 Db 61 GATTTGTTTAAATAACAATGATCAAGCAACTGAAATTTTAACTTCAGCAGCAAGACGTTA 120  
 Qy 121 GGTATGATTTACGTATCTCATCGATAG---GAGAGAGACAACTCAATCAGACCCGC 177  
 Db 121 GACTTTGATATTTTAGAGACAAATGTTTACTGATGAAGAGGTAAATTCGGGTGAACCTGAA 180  
 Qy 178 TATAGCAACACAGCCATTTCTAGCGACTTCGGTGTGATCTACCGTTTATTGCAAGAAAAG 237  
 Db 181 ACACGCAACAGCTTTTATGAGCGATAGTTTCGGCAATTA-----TTAGCAGCGCTAAA 234  
 Qy 238 GGTATACGCTGATATAGTTGCTGTTGCTCTTGGAGAAATCTCTGCTTGGTGCA 297  
 Db 235 ATTTTGAATCTGATTTTACTATGGGCAATAGTTTAGTGATATTTCAAGTTTGTGCA 294  
 Qy 298 AGCGGCGCTTGGATTTTGAAGATGCGGTTCGCTTGGTAGCTAAGCGTGGAGCCTATATG 357  
 Db 295 GCTGACGTATTATCATTTTGAAGATGCGATTTAAATTTGTAGAAACGTTGCTAATAG 354  
 Qy 358 GAAGAGCGGCTCTGCTGACTCTGGCAAGATGTTAGCAGTTCTCAATACCCAGTAGAG 417  
 Db 355 GCGCAAGCATTTCTCTACTGTTGAGAGCATGGCTGCAGTATTGGGATTTAGATTTTGAT 414  
 Qy 418 GTCATTGAAGAGCCTGTCA---AAAAGCTTCTGAACTTGGAGTGGTTACTCCAGCCAA 474  
 Db 415 AAAGTCGATGAATTTGTAAGTCAATTTATCTGATGACAAATAATTTGAACCCAGCAAC 474  
 Qy 475 TATAACACACCTGCACAAATCGTCAATTTCTGAGAGTGGTTGCGATTTGATCGAGCGTT 534  
 Db 475 ATTAATGCCAGGTCAATTTGTTTTCAGTTCACAAAGCTTTTAAATTTGATGAGCTAGTA 534  
 Qy 535 GAACTTTTGCAGAGCAGGTGCGCAACGCTTGAATTCCTCTTAAGGTCAGGTGCCCTTT 594  
 Db 535 GAAAAGGTAATCATTTAGTGCAGGCTGCTGCTGCTTTAGCAGTATCTGAGCCTTC 594  
 Qy 595 CACACCGCTCTCTTGGAGCGTGTAGCCAGAACTAGCTGAAACTCTAGCTCAGTAAAGT 654  
 Db 595 CATTCATCGCTAATGAAGATGATGAAGAGATTTTCAAGTTACATTAATCAATTTGAA 654  
 Qy 655 TTTTCAGATTTTACTTGTCCCTAGTGGGCAATACAGAGCT---GCTGTGATGCAAAA 711  
 Db 655 TGGCGTGAATGCTAAAGTTTCTCTGATTTCAAAATGTAATGCGCAAGTGAATGACAAA 714  
 Qy 712 GAGGACATTTGCTAGCTCTTGACGCGTCAGTCAAGGAAACCGTTTCTGTTCTATGAAAGT 771  
 Db 715 GAAGTAATTAATCTAATATGTTCAAGCAATATATTTACCAGTACATTAATCAATCACTCA 774

QY 772 ATTGGGCTCATGCAAGACGAGCATTAAGCAACTTTATTCGAGATTGGACCGGGGAAAGTC 831  
Db 775 ACAGAATGGCTAATAGACCAAGGTGTGATCATTTTATTGAATTTGGCTCGGAAAGTT 834  
QY 832 TTGTCAGGTTTGTAAAAAATTTGATCAAACTGCTCACTTAGCTCATGTGGAAGATCAA 891  
Db 835 TTATCTGGCTTAATTAATAAATAAATAGAGATGTTAAGTTAATCAATCAACTTTA 894  
QY 892 GCGAGTTTAGTACACTTTTAGAAAA 917  
Db 895 GAAGATGTGAAGGATGGAATGAAAA 920

## RESULT 7

US-09-090-793-13  
; Sequence 13, Application US/09090793  
; Patent No. 6140486  
; GENERAL INFORMATION:  
; APPLICANT: Calgene, LLC  
; TITLE OF INVENTION: Production of polyunsaturated fatty acids by expression  
; FILE REFERENCE: CGNE.131.01US  
; CURRENT APPLICATION NUMBER: US/09/090,793  
; EARLIER FILING DATE: 1998-06-04  
; EARLIER APPLICATION NUMBER: 60/048,650  
; EARLIER FILING DATE: 1997-06-04  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 13  
; LENGTH: 19227  
; TYPE: DNA  
; ORGANISM: Vibrio marinus  
US-09-090-793-13

Query Match 7.5%; Score 69.2; DB 3; Length 19227;  
Best Local Similarity 46.8%; Pred. No. 3.1e-12; Indels 15; Gaps 3;  
Matches 332; Conservative 0; Mismatches 363;  
QY 152 AAGAAGCAAACTCAATCAGACCGGCTATACGCAACCGCCATTCTAGCACTTCGGTTG 211  
Db 2401 AAGAAGCAATACGTTTAACGCAACATCGGCAACCGGATTGTTAGTGTG 2460  
QY 212 CTATCTACCGTTTATTCGAAGAAAGGGCTATCAGCCTGATATGTTGCTGTTGCTC 271  
Db 2461 GTCTGTTCAAAAGTTTAAAGCAAGAGGTTTAAAGCTGATTTGCTGCCGTCATAGTT 2520  
QY 272 TTGAGAAATCTCTGCTTGGTGCAGCGGCGCTTGGATTTCGAAGATCGGTTGCT 331  
Db 2521 TCGTGAGTTAACGCAATATGGCTGCCGATGATTGAGCGAAAGCGATTACATGATG 2580  
QY 332 TGGTAGCTAAGCTGGAGCCCTATATGGAAGAGCGGCTCTGCTGACTCTGGCAAGATGG 391  
Db 2581 TAGCGCTAGTCTGTTGTCAGCAATGGCTGCGCAGAGCAACAAGATTTTGAT-----G 2634  
QY 392 TAGCAGTTCTCAATACGCCAGTAGAGTCAATTCGAAGAAGCTCTCAAAAAGCTTCGAC 451  
Db 2635 CAGTAAGATGGCGCTGTTGTTGTTGATCCAAAGCAAGTCCGTGTGATCATGATACC 2694  
QY 452 TTGAGTGGTTATCCAGCCAACTATAACACACCTGCACAAATCGTCATTTGCTGGAGAAG 511  
Db 2695 TTGATGATGCTCTATTGCTAACTTCACTCGAATAACCAAGTTGTTATTGCTGGTACTA 2754  
QY 512 TGGTTGAGTTGATCGAGCGGTTGAATTTTTCGAAGAACAGGTGCGCAACGCTTGATTC 571  
Db 2755 CGGAGCAGGTTGTTAGCGGTTTACAACTTAGGTAATGCTGTTTCAAA---GTTGTC 2811  
QY 572 CTCTTAAGGTGTCAGTCCCTTTTACACCGCTCTCCTTGAGCCTGCTAGCCGAAACTAG 631  
Db 2812 CACTGCCGGTATCTGCTGCTTCCATAGACCTTTAGTTGCTCAGCGCAAAACCACTTG 2871  
QY 632 CTGAACTCTAGCTCAGGTAAGTTTTCAGATTTTACTGTTCCCTAGTCGGCAATACAG 691  
Db 2872 CTAAGCGGTTGATAGCGCTAAATTTTAAAGCGCAAGCATTCAGTGTTCGTAATGGCA 2931

QY 692 AAGCTGCTGTG-----ATGCAAAAAGAGACAACTTCTCAGCTCTTTGACCGCTCAGGTCA 745  
Db 2932 CAGGCTTGGTGCATTCAAGCAAAACCGAATGACATTAAAGAAAAACCTGAAAAACCATGTC 2991  
QY 746 AGGAACCGGTTCTGTTTCTATGAAGTATTGGGTCATGCAAGAAGCAGGCATAAGCAACT 805  
Db 2992 TGGAACTGTTTCAATTCATCAAGAATGACAACTATATGCTGATGTTGGCGCGGTAT 3051  
QY 806 TTATCGAGATTGACCGGGGAAAGTCTTGTACGTTTGTATAAAAAATT 855  
Db 3052 TTATCGAATTTGGTCCAAAGATGTATTAACTAAATTTGGTTGAAAAACATT 3101

## RESULT 8

US-09-090-793-12  
; Sequence 12, Application US/09090793  
; Patent No. 6140486  
; GENERAL INFORMATION:  
; APPLICANT: Calgene, LLC  
; TITLE OF INVENTION: Production of polyunsaturated fatty acids by expression  
; FILE REFERENCE: CGNE.131.01US  
; CURRENT APPLICATION NUMBER: US/09/090,793  
; EARLIER FILING DATE: 1998-06-04  
; EARLIER APPLICATION NUMBER: 60/048,650  
; EARLIER FILING DATE: 1997-06-04  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 12  
; LENGTH: 40138  
; TYPE: DNA  
; ORGANISM: Vibrio marinus  
US-09-090-793-12

Query Match 7.5%; Score 69.2; DB 3; Length 40138;  
Best Local Similarity 46.8%; Pred. No. 5e-12; Indels 15; Gaps 3;  
Matches 332; Conservative 0; Mismatches 363;  
QY 152 AAGAAGCAAACTCAATCAGACCGGCTATACGCAACCGCCATTCTAGCACTTCGGTTG 211  
Db 19384 AAGAAGCAATACGTTTAAACGCAACATCGGCAACCGGATTGTTAGTGTG 19443  
QY 212 CTATCTACCGTTTATTCGAAGAAAGGGCTATCAGCCTGATATGTTGCTGTTGCTC 271  
Db 19444 GTCTGTTCAAAAGCTTTAAAGCAAGAGGTTTAAAGCTGATTTGCTGCCGTCATAGTT 19503  
QY 272 TTGAGAAATCTCTGCTTGGTGCAGCGGCGCTTGGATTTCGAAGATCGGTTGCT 331  
Db 19504 TCGTGAGTTAACGCAATATGGCTGCCGATGATTGAGCGAAAGCGATTACATGATG 19563  
QY 332 TGGTAGCTAAGCTGGAGCCCTATATGGAAGAGCGGCTCTGCTGACTCTGGCAAGATGG 391  
Db 19564 TAGCGCTAGTCTGTTGTCAGCAATGGCTGCGCAGAGCAACAAGATTTTGAT-----G 19617  
QY 392 TAGCAGTTCTCAATACGCCAGTAGAGTCAATTCGAAGAAGCTCTCAAAAAGCTTCGAC 451  
Db 19618 CAGTAAGATGGCGCTGTTGTTGTTGATCCAAAGCAAGTCCGTGTGATCATGATACC 19677  
QY 452 TTGAGTGGTTTACTCCAGCCAACTATAACACACCTGCACAAATCGTCATTTGCTGGAGAAG 511  
Db 19678 TTGATGATGCTCTATTGCTAACTTCACTCGAATAACCAAGTTGTTATTGCTGGTACTA 19737  
QY 512 TGGTTGAGTTGATCGAGCGGTTGAATTTTTCGAAGAACAGGTGCGCAACGCTTGATTC 571  
Db 19738 CGGAGCAGGTTGCTGTTAGCGGTTTACAACTTAGGTAATGCTGTTTCAAA---GTTGTC 19794  
QY 572 CTCTTAAGGTGTCAGTCCCTTTTACACCGCTCTCCTTGAGCCTGCTAGCCGAAACTAG 631  
Db 19795 CACTGCCGGTATCTGCTGCTTCCATAGACCTTTAGTTGCTCAGCGCAAAACCACTTG 19854  
QY 632 CTGAACTCTAGCTCAGGTAAGTTTTCAGATTTTACTGTTCCCTAGTCGGCAATACAG 691

Db 19855 CTAAAGCGGTGATAGCGCTAAATTTAAAGCGCCAGCAATCCAGTGTGTTGCTAATGGCA 19914  
QY 692 AAGTGTGTG-----ATGCAAAAGAGGACATGCTGCTAGCTTTGACCGGTGAGTCA 745  
Db 19915 CAGCGTTGGTGCATTCAGCAACCGAATGACATTAAGAAAAACCTGAAAAAACCATGC 19974  
QY 746 AGGAACCGGTTCGTTTCTATGAAAGTATTGGGGTCATGCAAGAGCAGGCATTAAGCAACT 805  
Db 19975 TGGAACTGTTCATTTCAATCAAGAAATTGACACATCTATGCTGATGGTGGCGGGTAT 20034  
QY 806 TTATCGAGATGGACCGGGGAAATCTGTGACGTTTGTGTAATAAAATTT 855  
Db 20035 TTATCGAATTTGGTCCAAAGAATGATTAATACTAAATTTGGTTGAAACATT 20084

## RESULT 9

US-08-375-709-10  
; Sequence 10, Application US/08375709  
; Patent No. 5683898  
; GENERAL INFORMATION:  
; APPLICANT: YAZAWA, Kazunaga  
; APPLICANT: YAMADA, Akiko  
; APPLICANT: KATO, Seishi  
; APPLICANT: KONDO, Kiyosi  
; TITLE OF INVENTION: Gene Coding For Eicosapentaenoic Acid  
; TITLE OF INVENTION: Synthesizing Enzymes and Process for Production of  
; TITLE OF INVENTION: Eicosapentaenoic Acid  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/375,709  
; FILING DATE: 20-JAN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/178,251  
; FILING DATE: 14-MAY-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 4-147945  
; FILING DATE: 15-MAY-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: WEGNER, Harold C.  
; REGISTRATION NUMBER: 25,258  
; REFERENCE/DOCKET NUMBER: 53466/150/AAOK  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8268 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:  
; ORGANISM: Shewanella putrefaciens SCRC-2874 (FERM  
; ORGANISM: BP-1625)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..8268  
; FEATURE:  
; NAME/KEY: mat\_peptide

; LOCATION: 1..8268  
US-08-375-709-10  
Query Match 5.8%; Score 53.4; DB 1; Length 8268;  
Best Local Similarity 46.7%; Pred. No. 3.6e-07;  
Matches 260; Conservative 0; Mismatches 276; Indels 21; Gaps 2;  
QY 152 AAGAGAGCAAACTCAATCAGACCCGCTATACGCAACCAAGCCATCTAGCGACTTCGGTTG 211  
Db 1991 AAGAAAGCCATTTTGACCAATACCGCAATGCCAAAGCGCAATGGTGGATTTCAATGG 2050  
QY 212 CTATCTACCGTTTATTGCAAGAAAGGCTATACGCTGTATATGGTTCTGCTGTTTCTC 271  
Db 2051 GTCAATACGATTTGTTTACTGCGGCTGCTTTAATGCGACATGTTGCGAGCCATAGCT 2110  
QY 272 TTGGAGAACTACTCTGCTTGGTGGCAAGCGCGCTTGGATTTGAAGATGGGTTGCT 331  
Db 2111 TTGGTGAGCTAAGTCACTGTGCTGCAGGTGTTTATTTTCAGCTGATGACTACTACAAG 2170  
QY 332 TGGTAGCTAAGCGTGGAGCCTATATGGAAGAAAGCGGCTCCTGCT----- 375  
Db 2171 TGGCTTTTGTCTGCTGAGGCTATGCGCAACAAAAGCACCAGCTAAAGACGCGTTGAAG 2230  
QY 376 --GACTCTGGCAAGATGCTAGCAGTTCTCAATACCCAGTAGAGTCAATTGAAGAGCT 433  
Db 2231 CAGATGCGAGGCAATGTTTGCATATCAACCAAGAGTCTGCAGACCTTGAACCGTTG 2290  
QY 434 GTCAAAAAGCTTCTCAACTTGGAGTGGTTACTCCAGGCCAACTATAACACACCTGCACAAA 493  
Db 2291 AAGCCACCATCGCTAAATTTGATGGGTGAAAGTGCCTAATTAACGGCCCAACGCAAT 2350  
QY 494 TCGTCAATGCTGGAGAAAGTGGTTGCAATGATCGAGCGGTTGAACTTTTGAAGAGCAG 553  
Db 2351 CAGTAATTGCGAGGCCCAACAGCAACTACCGCTGATGCGGCTAAAGCGCTAACTGAGCTG 2410  
QY 554 GTGCCAAACGCTTGATCTCTTAAGGTGTCAGGTCCTTTTACACCCGCTCTCCCTTGAGC 613  
Db 2411 GTTACAAA---GCGATTAACTGCCAGTATCAGGTGCATTTCCACTGAATTTGTTGTC 2467  
QY 614 CTGTAGCGCAGAACTAGCTGAAACTCTAGCTCAGGTAAAGTTTTTTCAGATTTTACTTTC 673  
Db 2468 ACGCTCAAGCGCCATTTGCTAAAGCGGATTTGAGCGAGCCAAATTTTACTAAACCAAGCCGAG 2527  
QY 674 CCTAGTGGCAATACA 690  
Db 2528 CACTTTACTCAAAATGCA 2544  
RESULT 10  
US-08-752-929-10  
; Sequence 10, Application US/08752929  
; Patent No. 5798259  
; GENERAL INFORMATION:  
; APPLICANT: YAZAWA, Kazunaga  
; APPLICANT: YAMADA, Akiko  
; APPLICANT: KATO, Seishi  
; APPLICANT: KONDO, Kiyosi  
; TITLE OF INVENTION: Gene Coding for Eicosapentaenoic Acid Synthesizing  
; TITLE OF INVENTION: Enzymes and Process for Production of Eicosapentaenoic  
; TITLE OF INVENTION: Acid  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30

;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/752,929  
;; FILING DATE: 20-NOV-1996  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/375,709  
;; FILING DATE: 20-JAN-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/178,251  
;; FILING DATE: 14-MAY-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: JP 4-147945  
;; FILING DATE: 15-MAY-1992  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: WEGNER, Harold C.  
;; REGISTRATION NUMBER: 25,258  
;; REFERENCE/DOCKET NUMBER: 53466/150/AAOK  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (202)672-5300  
;; TELEFAX: (202)672-5399  
;; TELEX: 904136  
;; INFORMATION FOR SEQ ID NO: 10:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 8268 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; ORIGINAL SOURCE:  
;; ORGANISM: Shewanella putrefaciens SCRC-2874 (FERM  
;; ORGANISM: BP-1625)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 1..8268  
;; FEATURE:  
;; NAME/KEY: mat peptide  
;; LOCATION: 1..8268  
;;  
;; US-08-752-929-10  
;;  
Query Match  
Best Local Similarity 46.7%; Pred. No. 3.6e-07;  
Matches 260; Conservative 0; Mismatches 276; Indels 21; Gaps 2;  
5.8%; Score 53.4; DB 1; Length 8268;  
QY 152 AAGAGACAACTCAATCAGACCGCTATACGCAACCGCATTTCTAGGACITTCGGTTG 211  
DB 1991 AAGAGCCATTTTGCACCAATACCGCAATGCCAAAGCCCAATGGTGGCATTTCAATGG 2050  
QY 212 CTATCTACCGTTTATTGCAAGAAAGGGCTATCAGCGCTGATATGTTGTTGTTTGTCTC 271  
DB 2051 GTCAATACGATTGTTTACTCGCGCTGGCTTTAATGCCACATGTTGAGCGCCATAGCT 2110  
QY 272 TTGAGAAATACTCTGCGTTGGTGGCAAGCGCGCTTGGATTTTGAAGATGCGGTTGCGCT 331  
DB 2111 TTGGTGAGCTAAGTGACCTGTGTGCTGACGGTGTATTTCAGCTGATGACTACTACAAGC 2170  
QY 332 TGGTAGTAAGCGTGAGCGCTATATGCAAGAGCGGCTCTCTCTCTCTCTCTCTCTCTCTCT 375  
DB 2171 TGGCTTTTCTGCTGGTGGCTATGCAACCAAGCAGCGCTTAAGCGCGCTTGAAG 2230  
QY 376 --GACTCTGCAAGATGGTAGAGTTCTCAATACGCCAGTAGAGGTCAATTGAAGAGCGCT 433  
DB 2231 CAGATCAGGAGCAATGTTTGAATCATACCAAGAGTCTGCAGACCTTGAACCGTTG 2290  
QY 434 GTCACCAAGCTTCTGAACTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 493  
DB 2291 AAGCCACCATCGCTAAATTTGATGGGGTGAAGTGGCTTAATTAACGCCGCAACGCAAT 2350  
QY 494 TCGTCAATTGCTGGAGAAGTGGTTGCAAGTTGATCGAGCGGTTGAACTTTTGCAGAGAGCAG 553  
DB 2351 CAGTAATTGAGGCCCAACAGCAACTACCGCTGATGCGGCTAAAGCGCTAACTAGAGCTTG 2410  
QY 554 GTGCCAAAGCTGTGATCTCTTAAGGTGTAGGTCCTTTTACACCGCTCTCTCTTGGC 613  
DB 152 AAGAGACAACTCAATCAGACCGCTATACGCAACCGCATTTCTAGGACITTCGGTTG 211  
DB 15896 AAGAGCCATTTTGACCAATACCGCAATGCCCAATGGTGGCATTTCAATGG 15955

Db 2411 GTTACAAA--GCCATTAACTGCCAGTATCAGGTGCAATCCACACTGAACCTTGTGTC 2467  
QY 614 CTGCTAGCCAGAACTAGCTGAAACTCTAGCTCAGGTAAGTTTTCAGATTTTACTTGTGTC 673  
Db 2468 ACGCTCAAGCGCCATTTGCTTAAGCGATTGACGAGCCAAATTTACTAAAACAGCCGAG 2527  
QY 674 CCCTAGTCGGCAATACA 690  
Db 2528 CACTTTACTCAATGCA 2544

RESULT 11  
US-08-375-709-1  
; Sequence 1, Application US/08375709  
; Patent No. 5683898  
; GENERAL INFORMATION:  
; APPLICANT: YAZAWA, Kazunaga  
; APPLICANT: YAMADA, Akiko  
; APPLICANT: KATO, Seishi  
; APPLICANT: KONDO, Kiyosi  
; TITLE OF INVENTION: Gene Coding For Eicosapentaenoic Acid  
; TITLE OF INVENTION: Synthesizing Enzymes and Process for Production of  
; TITLE OF INVENTION: Eicosapentaenoic Acid  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/08/375,709  
; APPLICATION NUMBER: 375,709  
; FILING DATE: 20-JAN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/178,251  
; FILING DATE: 14-MAY-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 4-147945  
; FILING DATE: 15-MAY-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: WEGNER, Harold C.  
; REGISTRATION NUMBER: 25,258  
; REFERENCE/DOCKET NUMBER: 53466/150/AAOK  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 37895 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:  
; ORGANISM: Shewanella putrefaciens SCRC-2874 (FERM  
; ORGANISM: BP-1625)  
; US-08-375-709-1  
Query Match 5.8%; Score 53.4; DB 1; Length 37895;  
Best Local Similarity 46.7%; Pred. No. 9.3e-07;  
Matches 260; Conservative 0; Mismatches 276; Indels 21; Gaps 2;  
QY 152 AAGAGACAACTCAATCAGACCGCTATACGCAACCGCATTTCTAGGACITTCGGTTG 211  
DB 15896 AAGAGCCATTTTGACCAATACCGCAATGCCCAATGGTGGCATTTCAATGG 15955

QY 212 CTATCTACGCTTTATTCGAGAAAGGCGCTATACGCTGATATATGTTGCTGTTGCTC 271  
Db 15956 GTCAATACGATTTGTTTACTCGGCTGCTTTAATGCGCATGTTGCGAGCCATAGCT 16015  
QY 272 TTGGAGATACTCTGCTCTGTTGGCAAGCGGCTGATTTTGAAGATCGGTTGCT 331  
Db 16016 TTGGTACCTAAGTCACTGTGTGCTGAGTGTATTTTACGCTGATGACTACTACAAGC 16075  
QY 332 TGGTAGCTAAGCGGAGCCTATATGGAAGAGCGGCTCTGCT----- 375  
Db 16076 TGGCTTTTGTCTGCTGAGGCTATGGAACAAAGACCGCTAAAGACGGCTTGAAG 16135  
QY 376 --GACTCTGCGNAGTGTACAGTCTCAATACCGCAGTAGAGTCAATGAAGACCT 433  
Db 16136 CAGTGCAGGAGCAATGTTTGAATCATACCAAGAGTGTGCGAGCCTTGAACCTTG 16195  
QY 434 CTCAAAAGCTTCTGAATCTGGAGTGTACTCCAGCAACTATAACACACCTGCACAAA 493  
Db 16196 AAGCCACCATCGCTAAATTTGATGGGTGAAGTGCCTAACTATAACGCGCAACGAAT 16255  
QY 494 TCGTCAATGCTGGAGAGTGTGTGAGTGTATGAGCGGTTGAACCTTTGCAAGAACGAG 553  
Db 16256 CAGTAATTTGAGGCGCAACAGCAACTACCGCTGATGCGGCTAAAGCGCTAACTGAGTTG 16315  
QY 554 GTGCCAAAGCTTGTATCTCTTAAGGTGTGAGGTGTGAGCGGTTGAACCTTTGCAAGAACGAG 553  
Db 16316 GTTACAAA--GCGATTAACTGCGCAGTATCAGGTGCAATCCACACTGAACTTTGGTC 16372  
QY 614 CTGTAGCAGAACTAGCTGAACTCTAGCTCAGTAAAGTGTTCAGATTTTACTTTGTC 673  
Db 16373 AGCTCAAGCGCAATTTGCTAAAGCGATGACGAGCAAAATTTACTAAACAGCGGAG 16432  
QY 674 CCTAGTGGCAATACA 690  
Db 16433 CACTTTACTCAATGCA 16449

## RESULT 12

US-08-752-929-1  
; Sequence 1, Application US/08752929  
; Patent No. 5798259  
; GENERAL INFORMATION:  
; APPLICANT: YAZAWA, Kazunaga  
; APPLICANT: YAMADA, Akiko  
; APPLICANT: KATO, Seishi  
; APPLICANT: KONDO, Kiyosi  
; TITLE OF INVENTION: Gene Coding for Eicosapentaenoic Acid Synthesizing  
; TITLE OF INVENTION: Enzymes and Process for Production of Eicosapentaenoic  
; TITLE OF INVENTION: Acid  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: 20-NOV-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/375,709  
; FILING DATE: 20-JAN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/178,251  
; FILING DATE: 14-MAY-1993

;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: JP 4-147945  
;; FILING DATE: 15-MAY-1992  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: WEGNER, Harold C.  
;; REGISTRATION NUMBER: 25,258  
;; REFERENCE/DOCKET NUMBER: 53466/150/AAOK  
;; TELEPHONE: (202)672-5300  
;; TELEFAX: (202)672-5399  
;; TELEX: 904136  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 37895 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; ORIGINAL SOURCE:  
;; ORGANISM: Shewanella putrefaciens SCRC-2874 (FERM  
;; ORGANISM: BP-1625)  
;; US-08-752-929-1

Query Match 5.8%; Score 53.4; DB 1; Length 37895;  
Best Local Similarity 46.7%; Pred. No. 9.3e-07;  
Matches 260; Conservative 0; Mismatches 276; Indels 21; Gaps 2;  
QY 152 AAGAGACAACTCAATCAGACCCGCTATACGCAACCGCCATCTAGCGACTTCGGTTG 211  
Db 15896 AAGAGCCATTTTGACCAATACCGCAATGCCAAGCCCAATGGTGGATTTCAATGG 15955  
QY 212 CTATCTACCGTTTATTTGCAAGAAAGGGCTATCAGCCTGATATGTTGCTGTTGCTC 271  
Db 15956 GTCAATACGATTTGTTTACTGCGGCTGGCTTTAATGCGCAGATGTTGAGCCATAGCT 16015  
QY 272 TTGGAGAACTCTGCTTGGTGGCAAGCGGCGCTTGAATTTGAAGATCGGTTGCT 331  
Db 16016 TTGGTGAAGTAACTGCTGCTGCTGAGGTGTTTATTTACGCTGATGACTACTACAAGC 16075  
QY 332 TGGTAGCTAAGCGTGGAGCTATATGGAAGAGCGGCTCCTGCT----- 375  
Db 16076 TGGCTTTTGTCTGCTGAGGCTATGGCAACAAAGCACCGCTAAAGACGGCTTGAAG 16135  
QY 376 --GACTCTGCGAAGATGTTAGAGTCTCAATACGCCAGTAGAGTCAATGAAGACCT 433  
Db 16136 CAGATCAGGAGCAATGTTTGAATCATAAACCAAGAGTCTGCGACCTTGAACCGTTG 16195  
QY 434 GTCAAAAAGCTTCTGAACCTTGGAGTGTACTCCAGCCAACTATAACACACCTGCACAAA 493  
Db 16196 AAGCCACCATCGCTAAATTTGATGGGTGAAGTGTGCTAACTATACGCGCCCAACGCAAT 16255  
QY 494 TGTCTATTGCTGGAGAGTGTGTTGAGTGTGATCGAGCGGTTGAACCTTTTGAAGAGCAG 553  
Db 16256 CAGTAATTTGAGGCGCAACAGCAACTACCGCTGATGCGGCTAAAGCGCTAACTGAGCTTG 16315  
QY 554 GTGCCAAAGCTTGTATCTCTTAAGGTGTGAGGTGCTCCCTTTTACACCGCTCTCCTGAGC 613  
Db 16316 GTTACAAA--GCGATTAACTGCGCAGTATCAGGTGCAATCCACACTGAACTTTGGTC 16372  
QY 614 CTGTAGCAGAACTAGCTGAAACTCTAGCTCAGTAAAGTGTTCAGATTTTACTTTGTC 673  
Db 16373 AGCTCAAGCGCAATTTGCTAAAGCGATGACGAGCAAAATTTACTAAACAGCGGAG 16432  
QY 674 CCTAGTGGCAATACA 690  
Db 16433 CACTTTACTCAATGCA 16449

## RESULT 13

US-09-090-793-1  
; Sequence 1, Application US/09090793  
; Patent No. 6140486  
; GENERAL INFORMATION:

```

; APPLICANT: Calgene, LLC
; TITLE OF INVENTION: Production of polyunsaturated fatty acids by expression
; TITLE OF INVENTION: of polyketide-like synthesis genes in plants
; FILE REFERENCE: CGNE.131.01US
; CURRENT APPLICATION NUMBER: US/09/090,793
; CURRENT FILING DATE: 1998-06-04
; EARLIER APPLICATION NUMBER: 60/048,850
; EARLIER FILING DATE: 1997-06-04
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 37895
; TYPE: DNA
; ORGANISM: Shewanella putrefaciens
US-09-090-793-1

```

Query Match	5.8%;	Score 53.4;	DB 3;	Length 37895;
Best Local Similarity	46.7%;	Pred. No. 9.3e-07;		
Matches 260;	Conservative 0;	Mismatches 276;	Indels 21;	Gaps 2;

  

QY	152	AAGAAGACAAATCAATCAGACCCGGTATACGCAACACGACCATTCAGCGACTTCGGTTG	211
Db	15896	AAGAAGCCATTTGACCAATACCGCCAATGCCAAAGCGCAATTGGTGCATTCCAATGG	15955
QY	212	CTATCTACCGTTATTGCAAGAAAGGCGTATACGCGTGATATGTTTGCTGTTGTCTC	271
Db	15956	GTCAAACGATTTGTTTATCTGGCGGTGGCTTTAAAGCCACATGTTGACGCGCATAGCT	16015
QY	272	TTGGAGAATACTCTGCTTGGTGGCAAGCGGCGCTTGGAATTTTGAAGATGCGGTGGCT	331
Db	16016	TTGGTGAGCTAAGTGACCTGTGTGCTGCAGGTGTTATTTCACTGATGACTACTACAAGC	16075
QY	332	TGCTAGCTAAGCGTGAGCGCTATATGGAAGAAGCGCTCCTCTCT	375
Db	16076	TGCGTTTGGCTGTGTGAGGCTATGCGACAAAGACCGGCTAAAGCGCGTTGAAG	16135
QY	376	--GACTCTGCGAAGATGGTAGCAGTTCTCAATCGCCAGTAGAGGTCTATGGAAGAGCCT	433
Db	16136	CAGATCAGGAGCAATGTTTGCAATCATACCAAGAGTGCTGCAGACCTTGAAACCGTTG	16195
QY	434	GTCAAAAGGCTTCTGAACTTGGAGTGGTTACTCCAGCCAACTATAACACACCTGCACAAA	493
Db	16196	AAGCCACCATCGCTAAATTTGATGGGTGGAAGTCGCTAACTATTAACGGCCAAACGCAAT	16255
QY	494	TCGTCATTGCTGAGAAAGTGGTTGCAGTTGATCGAGCGGTTGAACTTTTGCAGAAGCAG	553
Db	16256	CAGTAATTGACGGCCCAACAGCAACTACCGCTGATCGCGCTAAACGGCTAACTGAGCTTG	16315
QY	554	GTGCCAAAGCCTTGATTCTCTTAAGGTCGAGTCCCTTTTCCACACCGCTCTCTCTTGAGC	613
Db	16316	GTTTACAAA--GCGATTAACTGCCAGTATCAGGTGCATTCACACTGAACCTTTGTTGTC	16372
QY	614	CTGCTAGCCAGAACTAGCTGAACCTCTAGCTCAGGTAGTATTTTCAGATTTTACTTGTCTC	673
Db	16373	ACGCTCAAGCGCCATTGCTTAAGCGAATTGACGACCCAAATTTTACTAAACCAAGCCGAG	16432
QY	674	CCCTAGTCGCGCAATACA	690
Db	16433	CACTTTACTCAATGCA	16449

RESULT 14  
US-09-103-840A-2  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; TITLE OF INVENTION: TUBERCULOSIS  
; FILE REFERENCE: 24366-20007.00

```

: CURRENT APPLICATION NUMBER: US/09/103,840A
:
: CURRENT FILING DATE: 1998-05-24
:
: NUMBER OF SEQ ID NOS: 2
:
: SOFTWARE: Patent Ver. 2.1
:
: SEQ ID NO 2
:
: LENGTH: 4403765
:
: TYPE: DNA
:
: ORGANISM: Mycobacterium tuberculosis
:
: FEATURE:
:
: OTHER INFORMATION: CDC 1551
:
: OTHER INFORMATION: "n" bases at various positions throughout the sequence
:
: OTHER INFORMATION: represent a, t, c or g
:
: US-03-103-840A-2

```

Query Match	5.8%;	Score 53.2;	DB 4;	Length 4403765;
Best Local Similarity	49.2%;	Pred. No. 2.2e-05;		
Matches 182;	Conservative 0;	Mismatches 173;	Indels 15;	Gaps 1;
QY	7	AAACAGAGCCTTTTATTTTGGCTGGTCAAGTGCACCGATATCTAGGGATGGACGGGATTC	66	
Db	489058	AGAAGCGCTTTTGCTTTCCCGCTCAGGGCTCCGACGTGGCTGGGTATGGTAGCAGCTT	489117	
QY	67	TATGATCAGTATCGGATTTCAAGAAACGATTGATCGACGAGTCAGGTGCTCGGTTAT	126	
Db	489118	TATGCTGCTACCGGTTTTCGCCGAGCGCCTCGATGCTGTGTCGACAGTTGGACCGG	489177	
QY	127	GATTTACGTTATC-----TCATCGATACGAAAGACAAACCTCAATCAG	171	
Db	489178	CACCTGCGGTATTCGCTCGCGGATGTGATCTGGGGGCACCCAGATCTGTTGAATACC	489237	
QY	172	ACCCGCTATACGCAACACGACCATTCAGCGACTTCGGTTGCTATCTACCGTTTATTGCAA	231	
Db	489238	ACCGAATTCGCCACGCGCGCTGTTTCGGTGGAGTGGCGCTGTATCGGCTGCTCATG	489297	
QY	232	GAAGAAGGGCTATCAGCCTGATATGTTGCTGTGGTTTGTCTTTGGAGAAATCTGCGCTTG	291	
Db	489298	TCGTGGGGGGTTCGGCCGGGTTTGCTGCTGGTCATTCGGTGGCGAGTTGGCCCGGGG	489357	
QY	292	GTGGCAAGCGGGCGCTTCGGATTTTGAAGATGCGGTTTGCCTTGGTAGCTAAGCGTGGAGCC	351	
Db	489358	CACGTCCGGGGCGCTGTGTTCCCGGATGCGCGGATGCTGTGTGGCCGCGGTGACCG	489417	
QY	352	TATATGGAAG	361	
Db	489418	TTGATGCAGG	489427	

```

RESULT 15
US-09-103-840A-1
> Sequence 1, Application US/09103840A
> Patent No. 6294328
> GENERAL INFORMATION:
> APPLICANT: FLEISCHMAN, Robert D.
> APPLICANT: WHITE, Owen R.
> APPLICANT: FRASER, Claire M.
> APPLICANT: VENTER, John C.
> TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
> TUBERCULOSIS
> FILE REFERENCE: 24366-20007.00
> CURRENT APPLICATION NUMBER: US/09/103,840A
> CURRENT FILING DATE: 1998-06-24
> NUMBER OF SEQ ID NOS: 2
> SOFTWARE: PatentIn ver. 2.1
> SEQ ID NO 1
> LENGTH: 441529
> TYPE: DNA
> ORGANISM: Mycobacterium tuberculosis
> OTHER INFORMATION: H37Rv
US-09-103-840A-1

```

Query Match 5.8%; Score 53.2; DB 4; Length 4411529;  
 Best Local Similarity 49.2%; Pred. No. 2.2e-05;  
 Matches 182; Conservative 0; Mismatches 173; Indels 15; Gaps 1;

Qy	7	AAAACAGCCTTTTATTTGCTGGTCAAGGTGCCCAAGTACTAGGAATGGACGGGATTC	66
Db	487616	AAGACGGCTTTTGTGTTTCCGGTCAGGGCTCGAGTGGTGGTATGGGTAGCGAGCTT	487675
Qy	67	TATGATCAGTATCCGATTGTCTAAAGAACGATTTGATCGACGAGTCAAGTCTCGGTTAT	126
Db	487676	TATGCTGCCATCCCGTTTTCGCCAGGCGCTTCGATGCTGTGTGGACGAGTTGGACCGG	487735
Qy	127	GATTTACGTTATC-----TCATCATACGGAAGACAAACTCAATCAG	171
Db	487736	CACCTGCGGTATCCGTCGGCGATGTGATCTGGGGGCACGACCAAGATCTGTTGATATCC	487795
Qy	172	ACCGGTATACGACACGACCAATCTACGCACTTCGGTGTCTATCTACCGTTTATTGCAA	231
Db	487796	ACCGAATTCGCCACGCGCGCGTGTTTCGGGTGGAGGTGGCGGTGATCGGCTGCTCATG	487855
Qy	232	GAAGAAGGGCTATCAGCGTGATATGTTGCTGGTTGTCTCTTGGAGAAATACTCTGCGTTG	291
Db	487856	TGTTGGGGGTGCGGCCGGGTTTGGTCTGGGTCAATTCGGTGGCGAGTTGGCCGGCGG	487915
Qy	292	GTGGCAAGCGGGCGCTTGGATTTTGAAGATCCGCTTGCTTGGTAGTAAAGCGTGGAGCC	351
Db	487916	CAGCTGCCCGGGGCGCTGTGTTCGCCGATGCGCGCATGCTGTGTGCCCGCGCTGGACGG	487975
Qy	352	TATATGGAAG	361
Db	487976	TTGATGCGAG	487985

Search completed: June 11, 2003, 19:47:06  
Job time : 88 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 11, 2003, 18:44:57 ; Search time 165 Seconds  
(without alignments)  
7811.963 Million cell updates/sec

Title: US-09-308-397-1  
Perfect score: 921  
Sequence: 1 atgacataaacagccctttt.....tagcacttttagaaaaaatg 921

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 870385 seqs, 69976893 residues

Total number of hits satisfying chosen parameters: 1740770

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTU5\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	911.4	99.0	921	10	US-09-815-242-9482
2	344.4	37.4	942	10	US-09-815-242-6823
3	340.4	37.0	3656	10	US-09-070-927A-577
4	204	22.1	950	10	US-09-974-300-1143
5	204	22.1	960	10	US-09-974-300-1086
6	196.4	21.3	939	10	US-09-815-242-6878
7	196.4	21.3	1830121	9	US-10-329-960-1
8	185.6	20.2	930	10	US-09-815-242-6028
9	158	17.2	6021	10	US-09-070-927A-458
10	139.6	15.2	592	10	US-09-974-300-5552
11	128	13.9	936	10	US-09-815-242-8032
12	127.6	13.9	906	10	US-09-815-242-4405
13	120.8	13.1	939	10	US-09-815-242-7796
14	118.4	12.9	6251	7	US-08-781-986A-25
15	69.2	7.5	7959	9	US-10-331-061-77
16	69.2	7.5	19227	9	US-10-331-061-13
17	69.2	7.5	40138	9	US-10-331-061-12
18	53.4	5.8	8268	9	US-10-331-061-83
19	53.4	5.8	37895	9	US-10-331-061-1

20	53.2	5.8	4209	9	US-09-712-363-20
21	51.8	5.6	64681	10	US-09-790-388-1
22	47	5.1	930	10	US-09-815-242-1793
23	40	4.3	268	10	US-09-815-242-1745
24	38.6	4.2	925	9	US-09-735-056-1
25	38.4	4.2	1236	9	US-10-077-584-3
26	38.2	4.1	1380	9	US-10-074-045-24
27	37.4	4.1	4851	9	US-09-712-363-116
28	37.2	4.0	532	9	US-10-184-644-564
29	37.2	4.0	532	9	US-10-184-644-564
30	36.6	4.0	520	9	US-10-184-644-332
31	36.6	4.0	520	9	US-10-184-644-332
32	36.6	4.0	15872	9	US-09-860-846-1
33	36.6	4.0	15872	9	US-09-988-384B-1
34	36.6	4.0	15872	10	US-09-836-821-1
35	36.6	4.0	15872	10	US-09-861-289-1
36	36.6	4.0	50937	9	US-09-808-880-1
37	35.6	3.9	3282	10	US-09-864-864-298
38	35.4	3.8	5558	10	US-09-070-327A-11
39	35.2	3.8	13842	9	US-09-860-846-30
40	35.2	3.8	13842	9	US-09-988-384B-30
41	35.2	3.8	13842	9	US-09-836-821-30
42	35.2	3.8	13842	10	US-09-861-289-30
43	35.2	3.8	36778	9	US-09-860-846-5
44	35.2	3.8	36778	9	US-09-836-821-5
45	35.2	3.8	36778	10	US-09-861-289-5

## ALIGNMENTS

RESULT 1  
US-09-815-242-9482  
; Sequence 9482, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haseibeck, Robert  
; APPLICANT: Chisen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9482  
; LENGTH: 921  
; TYPE: DNA  
; ORGANISM: Streptococcus pneumoniae  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(921)  
US-09-815-242-9482

Query Match 99.0%; Score 911.4; DB 10; Length 921;  
Best Local Similarity 99.3%; Pred. No. 8.9e-280;  
Matches 915; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ATGACTAAACAGCCTTTTATTTGCTGCTCAAGTGGCCAGTACTAGGAGTGGACGG 60  
Db 1 ATGACTAAACAGCCTTTTATTTGCTGCTCAAGTGGCCAGTACTAGGAGTGGACGG 60

QY 61 GATTTCATGATCAGTATCCGATTTGTCAAAGAAACGATTGATCGAGCAGTCAAGTGTCTC 120  
Db 61 GATTTCATGATCAGTATCCGATTTGTCAAAGAAACGATTGATCGAGCAGTCAAGTGTCTC 120

QY 121 GGTATGATTACGTTATCTCATGATCGGAGAGACAACTCAATCAGACCGCTAT 180  
Db 121 GGTATGATTACGTTATCTCATGATCGGAGAGAGACAACTCAATCAGACCGCTAT 180

QY 181 ACGCAACAGCCATTCTAGCGACTTCGGTTGCTATCTACCGTTTATTGCAAGAAAGGCG 240  
Db 181 ACGCAACAGCCATTCTAGCGACTTCGGTTGCTATCTACCGTTTATTGCAAGAAAGGCG 240

QY 241 TATCAGCCTGATATGGTTGCTGGTTGCTCTTGGAGATATCTGCTTGGTGGCAAGC 300  
Db 241 TATCAGCCTGATATGGTTGCTGGTTGCTCTTGGAGATATCTGCTTGGTGGCAAGC 300

QY 301 GCGCGCTTGGATTTGAAGATCGGTTGCTTTGGTAGCTTAAGCGTGGAGCCTATATGAA 360  
Db 301 GCGCGCTTGGATTTGAGATCGGTTGCTTTGGTAGCTTAAGCGTGGAGCCTATATGAA 360

QY 361 GAAGCGGCTCTGACTCTGGCAAGATGTTAGCAGTTCTCAATACGCCAGTAGAGTGC 420  
Db 361 GAAGCGGCTCTGACTCTGGCAAGATGTTAGCAGTTCTCAATACGCCAGTAGAGTGC 420

QY 421 ATTGAAGAGCGCTGTCAAAAGAGCTTCTGAACCTGGAGTGGTTACTCCAGCCAACTATAAC 480  
Db 421 ATTGAAGAGCGCTGTCAAAAGAGCTTCTGAACCTGGAGTGGTTACTCCAGCCAACTATAAC 480

QY 481 ACACCTGCACAAATGCTATCTGAGAGAGTGGTTGCAAGTGTATCGAGCGGTGAACTT 540  
Db 481 ACACCTGCACAAATGCTATCTGAGAGAGTGGTTGCAAGTGTATCGAGCGGTGAACTT 540

QY 541 TTGCAAGAGAGCGGTGCCAAAGCTTGATTCCTCTTAAGGTCTCAGGTCCCTTTCACACC 600  
Db 541 TTGCAAGAGAGCGGTGCCAAAGCTTGATTCCTCTTAAGGTCTCAGGTCCCTTTCACACC 600

QY 601 GCTCTCTTTGAGCCTGCTAGCCAGAACTAGCTGAAACTCTAGCTCAGGTAACTTTTCA 660  
Db 601 GCTCTCTTTGAGCCTGCTAGCCAGAACTAGCTGAAACTCTAGCTCAGGTAACTTTTCA 660

QY 661 GATTTTACTTGTCCCTAGTGGCAATACAGAGCTGCTGTGATGCAAAAAGAGACATT 720  
Db 661 GATTTTACTTGTCCCTAGTGGCAATACAGAGCTGCTGTGATGCAAAAAGAGACATT 720

QY 721 GCTCAGCTCTGACGCGTCAGGTCAAGGAACCGTTCTGTTCTATGAAGTATTGGGGTC 780  
Db 721 GCTCAGCTCTGACGCGTCAGGTCAAGGAACCGTTCTGTTCTATGAAGTATTGGGGTC 780

QY 781 ATGCAAGAGAGCGCATAGCAACTTTATCGAGATGGACCGGGGAAAGTCTGTTCAGGT 840  
Db 781 ATGCAAGAGAGCGCATAGCAACTTTATCGAGATGGACCGGGGAAAGTCTGTTCAGGT 840

QY 841 TTTGTTAAAAAATTGATCAAACTGCTCACTTGTAGTCTCATGTGGAAGATCAAGCGAGTTA 900  
Db 841 TTTGTTAAAAAATTGATCGAAGTCTCACTTGTAGTCTCATGTGGAAGATCAAGCGAGTTA 900

QY 901 GTAGCACTTTTGAAGAAATAG 921  
Db 901 GTAGCACTTTTGAAGAAATAG 921

## RESULT 2

US-09-815-242-6823

; Sequence 6823, Application US/09815242

Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Chisen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA 011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
PRIOR FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 6823  
LENGTH: 942  
TYPE: DNA  
ORGANISM: Enterococcus faecalis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(942)  
US-09-815-242-6823

Query Match 37.4%; Score 344.4; DB 10; Length 942;

Best Local Similarity 61.3%; Pred. No. 4.5e-99;

Matches 555; Conservative 0; Mismatches 351; Indels 0; Gaps 0;

QY 7 AAAACAGCCTTTTATTTGCTGCTCAAGTGGCCAGTACTTAGGGATGGACGGGATTC 66  
Db 16 AAAACAGCCTTTTATTTAGTGGACAAAGAGCCAGTATCAAGGGATGGTGAAGAATTA 75

QY 67 TATGATCAGTATCCGATTGTCAAAGAAACGATTGATCGAGCGAGTCAGTGTCTGGTTAT 126  
Db 76 TATCACCAGAAGCAGTATTGTCGGGAACTTTCGATGAAGCAAGTCTATCTTAGTTAT 135

QY 127 GATTTACGTTATCTCATCGATAGCGAAGAACAACTCAATCAGACCCGCTATACGCAA 186  
Db 136 GAGATGCGAGAACTTTGTTTACTGAAATGAACGTTTAAATGAAACAGATATACGCAA 195

QY 187 CCAGCCATCTAGCAGTTCGGTTGCTATCTACCGTTTATTCGAGAAAGGGCTATCAG 246  
Db 196 CCGTCTATTTTAAACAGTCAAGTTCGCAATTTTACCGCTTTTTCACAAAAGGACTAAACG 255

QY 247 CCGTATATGTTGCTGTTGCTCTCTTGGAGAACTACTCTGCGCTTGGTGGCAAGCGCGCC 306  
Db 256 CCGTATGCTGAGCGGTTTAACTTAGGGATATAGCGCTTGGTCCAGCGGGCT 315

QY 307 TTGGATTTGAAGATCGGTTGCTTGGTAGCTAGCGTGGAGCCCTATATGGAAGAGCG 366  
Db 316 TTGCGCTTTTTCAGAGCAGTGGCCTTTGGTCCAAAAGCGCGGTCACTACATGACAGAAGA 375

QY 367 GCTCCTGCTGACTCTCGCAAGATGGTAGCAGTTCTCAATACGCCAGTAGAGGTCAATTGAA 426  
Db 376 GCACCAAGAGAACTGGCAAAATGGTTGCTGTCATGAACTGCTGAGCGTGAAGTAATTGAG 435

QY 427 GAAGCCTGTCAAAAAGCTTCTGAACCTTGGAGTGGTTACTCCAGCCAACTATAACACACT 486

```

436 AAAGCCTGCCAGAGCCAGTCTTTTCGGAATTTGGCTCCAGCAATTTATATACACCA 495
487 GCACAAATCGTCAATTCGTGGAGAGTGGTTCGAGTTCATCGAGCGGTGGAACCTTTTGCAA 546
496 CAACAAATCGTGAATGGTGGTGGTGGTCTGCTGTTGATCAAGCGATGACACTTCTCAA 555
547 GAACAGGTGCCAAACGCTTGATTCCTCTTAAGGTGTCAGTCCCTTTACACCGCTCTC 606
556 GAACCTGGTGAAGCGAATGATTCCTTAATGTAGTGGCCCTTTCCATACGGCGTG 615
607 CTTCAGCCTGCTAGCCAGAACTAGCTGAAACTCTAGCTCAGTAACTTTTTCAGATTTT 666
616 TTACACAGCATCAAAAAAATTTGGCTCAGGATTTAGCAAAATTTGAACCTTCAACGATG 675
667 ACTTGTCCTAGTCGGCAATACGAAGCTGCTGTGATGCAAAAGAGGACATGCTCTAG 726
676 CAATTCCTGCTAATAGTAATAGCTGCGCAAAATTTATGCCCAAGAGCAATTTCAAGCG 735
727 CTCTTGACGGCTCAGGTCAAGGAACCGTTTCGTTTCTATGAAAGTATTGGGGTCATGCAA 786
736 TTATGGAAGCAAGTCATGCTCGGTAGCTTTTGAAGACAGTATCGAAACGATGAAG 795
787 GAACAGCGATAGCAACTTTATGAGATTGGAACCGGGAAAGTCTTGTGAGGTTTGT 846
796 GCTATGAACGTGGGAACGATGATGAAGTTGGTCCAGGGGAAACATTAACCTGTTTGT 855
847 AAAAAAATTCATCAACTGCTCACTTAGCTCATGTGAGATCAAGCGGTTTAGTAGCA 906
856 AAAAAAATTCACAAACAATGAAATGCAACCGTGGTGAAGATTTTGAACATTTAAGCAA 915
907 CTTTTTA 912
916 ACGTTA 921

```

RESULT 3

```

US-09-070-927A-577
; Sequence 577, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
;           Patrick J. Dillon
;           Steven Barash
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 577:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3656 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 577:
US-09-070-927A-577

Query Match      37.0%; Score 340.4; DB 10; Length 3656;
Best Local Similarity 61.7%; Pred. No. 2e-97;
Matches 556; Conservative 1; Mismatches 343; Indels 1; Gaps 1;

QY 7 AAACAGCGCTTTTATTTTGTGTCGAGTCCAGGTCCAGTATCTAGGATGGGAGGATTTTC 66
DB 2753 AAACAGCGGATTTTATTTAGTGGCAAGGAGCCAGTATCAAGGATGGGTGAAGATTA 2812
QY 67 TATGATCAGTATCCGATTTCAAGAAACGATTTGATCGAGCGAGTCAGGTGCTCGGTTAT 126
DB 2813 TATCACCAGGAGCGATTTGTTCCGGAACTTTTCGATGAAGCAAGTCATATCTTAGGTTAT 2872
QY 127 GATTTACGTTTATCTCATCGATACGGAAGAACAACTCAATCAGACCCGCTATACGCAA 186
DB 2873 GAGATGGCAGAACTTTGTTTACTGAAATGAACTTAAATGAAACAGATATACGCAA 2932
QY 187 CCAGCATTCTAGGACCTTCGTTGCTATCTACCGTTTATTCGAGAAAGAGGCTATCAG 246
DB 2933 CTTGTTATTTTAAACAGTCAGTGTGCGATTTTACCGTCTTTTGCACAAAAAGGACTAACG 2992
QY 247 CTTGATATGTTGCTGTTGTTGTTCTTTGGAGAACTCTGCTTGGTGGGAGCGGCGCC 306
DB 2993 CTTGATGCTGAGCGGTTTAAAGCTTAGGGAAATACAGTGTCTTTGGTCCAGCGGGGCT 3052
QY 307 TTGGATTTTGAAGATGCGGTTGCTTGGTAGCTAAGCGTGGAGCCTATATGGAAGAGCG 366
DB 3053 TTGGCTTTTTCAGAAAGCAGTGGCTTGGTCCAAAGCGCGTCTAGTACATGACAGAGCA 3112
QY 367 GCTCTGCTGACTCTGGCAGATGCTAGCAGTCTTCAATACGCCAGTAGAGTCAATGAA 426
DB 3113 GCACCAAGGAACTGGCAAAATGGTGTCTTTATGAAATGTGAGCGTGAAGTAATTCAG 3172
QY 427 GAAGCCTGTCAAAAGCTTCTGAACTTGGAGTGTACTCTCAGCCAACTATAACACACCT 486
DB 3173 AAAGCTGCCAAGAGCAGTGTCTTCGGAATTCGTGCTCCAGCAATTTATAATACACCA 3232
QY 487 GCACAAATCGTCAATTCGTGGAGAAAGTGGTTCAGTTCATCGAGCGGTGAACTTTGCAA 546
DB 3233 CAACAAATCGTGAATTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3292
QY 547 GAAGCAGGTGCCAAACGCTTCTTAAAGTGTGAGTCCCTTTTACACCGCTCTC 606
DB 3293 GAAGCTGGTGGAGAGCAATGATTCGGTAAATGTGAGTGGCCCTTTCCATACGCGCTG 3352
QY 607 CTTGAGCCTGTAGCCAGAAACTAGCTGAAACTCTAGCTCAGGTAAGTGTTCAGATTTT 666
DB 3353 TTACAAACCCAGCATCAAAAAAATTTGGCTCAGGATTTAGCAAAATTTGAACTTTCAAC 3412
QY 667 ACTTGTCCCTAGTCGGCAATACAGAGCTGCTGTGATGCAAAAGAGGAGGACTGCTCAG 726
DB 3413 CAATTTCTGTCATTAGTATACGACTGCGCAAAATTTATGCCCAAGAGGCAATTTCAAGCG 3472
QY 727 CTTTGAAGCGCTCAGGTCAAGGAAACCGGTTCTGTTTCTTATGAAAGTATTTGGGGTCATG 786
DB 3473 TTAATGGAAAAAGCAAGTCATGCTCGCGTACGTTTTCGAGACAGTATCGAAACGATGAAG 3532
QY 787 GAACGAGGCATAGCAACTTTTATCG-AGATTGGACCGGGGAAAGTCTTGTGAGGTTTTGT 845
DB 3533 GCTATGAACGTAGGAACGATGATTTGAAAGTTNGTCCAGGGGAAACATTAACYGGTTTTGT 3592
QY 846 TAAAAAAATTTGATCAAACTGCTCACTTAGCTCATGTGGAAGATCAAGCGAGTTTAGTAGC 905

```



```

Db 301 GCGGCGCTTGTGCTTAAAGATCGGTTATGCCGTGAGAAGCGCGCGAATCATG 360
Qy 358 GAAGAAGCGGCTCTGTGACTCTGGCAAGATGATAGCAGTTCTCAATAGCCAGTAGAG 417
Db 361 AATGAAGCGGCTGCGGCGGAGAGGCGGATGCGGCCATTCTCGCATGGACAGCCAG 420
Qy 418 GTCAATTCAGAAGCGCTCTCAAAAGCTTCTGAAGTTGGA---GTGGTTACTCCAGCAAC 474
Db 421 GCGCTGAAGAAGTGACGCGCAAAATTTCCGAAGAGGAAACCTTGTTCAGCTCGCCAAT 480
Qy 475 TATAACACACCTGCACAAATCGTCATTGCTGGAGAAGTGTTCGACAGTTGATCGAGCGGTT 534
Db 481 TTGAAGTGGCTGGGCAATCGTCATCTCGGAACAGCTTAAGCGGTGAGCTCGCTTCA 540
Qy 535 GAACCTTTTGAAGAGCAGGTGCGCAACGCTTGATCTCTTAAGGTGTCAGTCCCTTT 594
Db 541 GAGCTTCGAAGAAGAAAGGCGCAAAACGCGCGGATTTCTCTCGAAGTCAGCGGCGGTT 600
Qy 595 CACACCGCTCTCTCTTGAAGCTGCTAGCCAGAAACTAGCTGAAACTCTAGCTCAGGTAAGT 654
Db 601 CATTCGAGCTGATGAAGCGGAGCTGATAGCTTCGTGAAGTTCTTGAATGCTGCAG 660
Qy 655 TTTTCAGATTTACTTGTCCCTAGTGGCAA---TACAGAAGCTGCTGTGATGCAAAA 711
Db 661 ATCAACGACGAGCAATTCGCGTCTCCAAAGTAAACGCGGACCTTTGTACGGATAAA 720
Qy 712 GAGGACATTTACTTGTCCCTAGTGGCGCTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTC 771
Db 721 GACGACATTAAGAATAAAGTATGAACAGCTGTATTCCTCTGACGCTTTGAAGAAACA 780
Qy 772 ATTGGGTCATGCAAGAGCAGCATAAGCAACTTTATCGAGATTGGACCGGGAAGTC 831
Db 781 ATCAGCGCGCTGATTGCAAGAGCGCTCAGCAGCTTCATTGAAATCGTCCCGGAAAGTT 840
Qy 832 TTGTCAGGTTTGTAAAAAATTGATC 859
Db 841 TTGTCAGGCTTGTGAAGAAGTGAACC 868

```

RESULT 6

```

US-09-815-242-6878
; Sequence 6878, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITEA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110

```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6878
; LENGTH: 939
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(939)
US-09-815-242-6878

```

```

Query Match 21.3%; Score 196.4; DB 10; Length 939;
Best Local Similarity 54.1%; Pred. No. 6.7e-52;
Matches 470; Conservative 0; Mismatches 386; Indels 12; Gaps 3;

Qy 1 ATGACTAAACAGCCCTTTTATTTGCTGCTCAAGTGCCAGTATCTAGGATGGAGCG 60
Db 1 ATGAAAAATTCGAATGGTCTTCCAGTCAAGGCTCCCAAACTGTGCTGTTGCT 60
Qy 61 GATTTCTATGATCATGATCCGATTCGCAAGAAAGATGATGAGCGAGTCAGGTGCT 120
Db 61 GATCTTGAACACTGATATCCAAATCGTTATTCGAAACATTTAAACAAGCATCTGATGCGTT 120
Qy 121 GGTATGATTTACGTATCTCATCGATACGG---AAGAAGACAACTCAATCAGACCCGC 177
Db 121 GGTATGATTTATGATATCTTTTCAACAGGTCCAGTGAAGAACTTAATAAAACTTGG 180
Qy 178 TATAGCAACAGCAGCATTTAGCGATTCGTTGCTTCTCTTTGGAGAACTCTGCGCTTG 237
Db 181 CAACTCAGCCCGACATTTTATGCTGCTTATGCTGCTATTTATCGCTATGGAAGAAAA 240
Qy 238 GGTATC-----AGCCTGATATGTTGCTGCTTCTCTTTGGAGAACTCTGCGCTTG 291
Db 241 TTTCTCAATTAATAAACAGAAAGTATGCGAGTCAATAGCTTAGTGAGTATTTCTGCGTTA 300
Qy 292 GTGCAAGCGCGCTTGGATTTTGAAGATCGGTTGCTTGGTGTAGCTAAGCGTGGAGCC 351
Db 301 GTTGTGCTGCGGTGTTGATTTCCAAAGTCCGATTAATTTAGTGAATTCGCGGAAAA 360
Qy 352 TATATGAAGAAGCGGCTCTGCTGACTTCGGCAAGATGCTAGCAGTTCTCAATACGCCA 411
Db 361 TTAATGCAACAGCTGTGCTGAGGCACTGCGCAATGTATGCAATCATTTGTTAGAT 420
Qy 412 GTAGAGTCATGAAGAAGCGCTGCAAAAGCTTCTGAACTTGGAGTGTGTACTCCAGCC 471
Db 421 AATGAAGCAATTTATTAATGCTTTCGAAACAAAGCAGAGGAAGCGAAGTCGTATCTGCGGTG 480
Qy 472 AACTATACACACCTGCAACAACTGTCATTCGAGAGAGTGTGAGTTGATCGAGCG 531
Db 481 AACTTTAACTCAGCGGTCAAGTAGTTATTCGCGGTGCGAAGCTGAGTTGAGCGTGC 540
Qy 532 GTTGAACCTTTTGAAGAAGCAGGTGCCAAAGCTTTGATTCCTTTAAGGTGTGAGGTCCC 591
Db 541 GCTGCATTTATGTAAGAAGCAGGCGGCAACGTCATTCGCTGAGCTGAGCGTACCT 600
Qy 592 TTTCAACCGCTCTCTTGTAGCCCTGCTAGCCAGAACTAGCTGAACCTAGCTCAGGT- 650
Db 601 TCTCACTGTGCATTAATGAACCTGACGCGAGCAATTAGCGGTAAACACTTGAAGATTT 660
Qy 651 --AAGTTTTTTCAGATTTTACTTTGCTCCCTAGTCGGCAATACAGAAGCTGTGTGATGCAA 708
Db 661 CAAATTAATACACCAACAAATATCGGTATTAATAACCTTGATGTGAAGCTGAACTGAA 720
Qy 709 AAGAGGACATTTGCTCAGCTTTGACCGGTGCTAGGCAAGGAAACCCGTTGTTCTATGAA 768
Db 721 GGCACCGAAATTCGTACCGCACTTGTGCTGCTATATAGTCAGTTGCTGAGTGAAGT 780
Qy 769 AGTATTGGGTCATGCAAGAAGCAGGATAAGCAACTTTTATCGAGATTGACCGGGGAAA 828
Db 781 ACAGTTGAAGAAATGGCGCAGATGCGGTCTAGTGTCTGCTGAAGTGGCGCCAGGTA 840
Qy 829 GTCCTGTCAGGTTTGTAAAAAATTG 856
Db 841 GTATTAATGTTTAAACCAACGCAATTG 868

```

```
RESULT 7
US-10-329-960-1/c
; Sequence 1, Application US/10329960
; Publication No. US20030099277A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag
; TITLE OF INVENTION: Theorof, and Uses Thereof
; FILE REFERENCE: PB186P1
; CURRENT APPLICATION NUMBER: US/10/329,960
; CURRENT FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (40808)..(40810)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44416)..(44416)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44905)..(44905)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45593)..(45593)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45732)..(45732)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (47036)..(47036)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51334)..(51334)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51602)..(51602)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51786)..(51786)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51805)..(51805)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (55369)..(55369)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65309)..(65309)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65313)..(65313)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (80024)..(80024)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (100091)..(100091)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (102696)..(102696)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (105121)..(105121)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (107248)..(107248)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (117136)..(117136)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (119750)..(119750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (119924)..(119924)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (120038)..(120038)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
```





```
QY 579 GGTCTAGTCCCTTTTCACACGGCTCTCCCTTGAGCCTGCTAGCCAGAAACTAGCTGAAC 638
DB 292 TGTAGTGGCCCTTTTCCATACGGGCTGTTACACACGAGCATCAAAAAAATTGGCTCAGGA 351
QY 639 TCTAGCTCAGTAAGTTTTTCAGATTTTACTTGTGCCCTAGTGGCAATACAGAAGCTGC 698
DB 352 TTTAGCAAAATTGACATTTCAACAGATGCAAAATTCCTGTATAGTATAGTACTGCCGA 411
QY 699 TGTGATGCAAAAAGAGACATTTGTCTAGCTCTTGACCGCTCAGGTCAAGAAACCCGTTG 758
DB 412 AATTATGCCCAAGAGGCAATTAAGCGTTATTGAAAAAGCAAGTCATGTCTCGGTACG 471
QY 759 TTTCTATGAAAGTATTGGGTCATGCAAGAGCAGGAGGATAGCAACTTTATCCGATTCG 818
DB 472 TTTTGAAGACAGTATCGAAGCATGAAAGCTATGAACTGAGGACGATGATTTGAAGTTGG 531
QY 819 ACCGGGGAAGCTTTGTCTAGCTTTTGTATAAAAAATTGATCAAACTGCTCACTTAGCTCA 878
DB 532 TCCAGGGAACATTAATCTGTTTGTATAAAAAATTGACAAAAAATTTGAATGACCG 591
QY 879 TGTGAAGATCAAGCGAGTTTAGTACACTTTTA 912
DB 592 TGTGAAGATTTGCAACATTAACAGAAACGTTA 625

RESULT 10
US-09-974-300-5552
; Sequence 5552, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: Expression
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5552
; LENGTH: 592
; TYPE: DNA
; ORGANISM: Bacillus clausii
US-09-974-300-5552

Query Match 15.2%; Score 139.6; DB 10; Length 592;
Best Local Similarity 54.8%; Pred. No. 6.3e-34;
Matches 321; Conservative 0; Mismatches 259; Indels 6; Gaps 2;

QY 2 TGAGTAAACAGCCCTTTTATTCCTGTCGAAGTGCCCAAGTATCTAGGGATGGGCGG 61
DB 2 TGGGGAAGTTGCCCTTTTATTCCTGCAAGGCTCAAAAAGTTGGATGGGAGCAA 61
QY 62 ATTTCTATGATCATGATCCGATTCGAAAGAAAGATGATCGAGCGAGTCAGTCTCG 121
DB 62 CGCTTTTGTGTAAGATTTGATGTCAAAAGGAAGTGTGAGCGGCTGACCAAGCGTTG 121
QY 122 GTTATGATTTACGTTATCTCATCG--ATACGGNAGAGCAAACTCAATCAGACCGCT 178
DB 122 GTTTTCCTCTTCAGAGATCATCGAACATGACCCAGAGACAAGCTGAAACAAACGGCT 181
QY 179 ATAGCGCAACCAAGCACTTCTAGCGACTTCGGTTGCTATCTACCTGTTTATTCGAAAGAAAG 238
DB 182 ATGCCCAACCTGCTTTAGTGACGATGAGCACTGCAAGTCTCTCAATGTTTCGGGACGCTG 241
QY 239 GCTATACGCTGATATGGTTGCTGGTTTGTCTCTTGAGAAATACTCTGCTCTGGTGCAA 298
DB 242 ACATTCAGCAGATTTTGTGTCAGGCCATAGCCTAGGGGAGTATTCGGCTTTAGTGGCAT 301
```

```
QY 299 GCGGCGCCTTGATTTTGAAGATGCGGTTGCTTGGTAGCTAAAGCGTGGAGCCTATATGG 358
DB 302 GTAAGTCCCTTACATTTTGAAGATGCGGTCACACTTGTTCATCAACGGGGACACTCATGG 361
QY 359 AAGAAGCGGCTCCTGCTGACTCTGCAAGATGGTAGCAGTTCTCAATACGCCAGTAGAGG 418
DB 362 AGAAGCTGTGTGCCAAGACAGCAAGGGGGATGGCAGCGTACTCGGCTTAATAAAGAGG 421
QY 419 TCATTTGAAGAA---GCCTGTCAAAAAGCTTCTGAACITTTGGAGTGGTTACTCCAGCCAACT 475
DB 422 AGCTTGAGGAAGTGGCTTCCGAAATTTGACGGGACGGCGAAGTTGCTGAACTTGCCAATT 481
QY 476 ATAAACACACCTGCACAAATCGTCAATGCTGGAGAGTGGTTGCAAGTTCATCGAGCGGTTG 535
DB 482 TGAATTTGCCAGGGGCAAAATTTGCTGTGTCAGGACATGCGAAAGGAATTTGAAACAGCAGG 541
QY 536 AACTTTTGAAGAACAGGTCGCAAAACGCTTGATTCCTCTCTTAAGGT 581
DB 542 TATTAGCAAAACAAAAGGGGCCAAGCGTGTCTTGCCACTTGCCGT 587

RESULT 11
US-09-815-242-8032
; Sequence 8032, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: Prokaryotes
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8032
; LENGTH: 936
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(936)
US-09-815-242-8032

Query Match 13.9%; Score 128; DB 10; Length 936;
Best Local Similarity 49.2%; Pred. No. 4.2e-30;
Matches 456; Conservative 0; Mismatches 455; Indels 15; Gaps 4;

QY 1 ATGCTAAACAGCCCTTTTATTCCTGTCGAAGTGCCCAAGTATCTAGGATGGGCGG 60
DB 10 ATGATAAACAGCAATTTATTTCCGGACAAGGTGCCCAAAAGTTGGTATGGCAA 69
```

Qy	61	GATTCTCTATGATCAGTATCCGATTTGTCAAAGAAACGATTTGATCGAGCGAGTCTCAGGTGCTC	120
Db	70	GATTTTATAAACAACAACGATCAAGCAACTGAAATTTTAACTTCAGCGACGAAACAGGTTA	129
Qy	121	GGTTATTGATTTAGCTTATCTCATCGATACGGAGAGAGA--CAAACTCAATCAGACCCGC	177
Db	130	GACTTTGATATTTTAGAGACAATGTTTACTGACGAGACGGCAAAATTTGGGTGAAACTGAA	189
Qy	178	TATACGCAACACGACCATTTAGCGACTTCGGTTCGTATCTACCGGTTTATTTCGAAGAAAAAG	237
Db	190	AACACGCAACACGCTTTATTGACGCATAGCTCGGGGTTA-----TTAGCGGGCTAAAA	243
Qy	238	GGCTATCAGCCTGATAGTGGTCTGGTTCCTCTCGAGAACTACTCGCTTGGTGGCA	297
Db	244	AAATTTGAATCCCGATTTTACTATGGGGCATAGTTTAGTGAATATTCAAGTTTAGTTGCA	303
Qy	298	AGCGGGCCCTTGGATTTTGAAGATCGGTTGCCCTTGGTAGCTAGACGTGGAGCCTATATG	357
Db	304	GCTGACGTATTATCAATTTGAAGATCGAGTTAAATTTGTAGAAAACGTGCTCAATTAATG	363
Qy	358	GAAGAAGCGGCTCTGCTGACTCTCGGCAAGATGTTAGCGAGTTCTCAATACGCCAGTAGAG	417
Db	364	GCCAAAGCATTTTCTACTGGTGTAGGAAGCATGGCTGCAGTATTGGGCTTTAGATTTTGAT	423
Qy	418	GTCAATTTGAAGAAAGCCTGTCA---AAAAGCTTCTGAACCTGGAGTGGTTACTCCAGCGAAC	474
Db	424	AAAGTCGATGAAATTTGTAACTCATCTCATGACAAAATAATTGAACACGACGAAC	483
Qy	475	TATAACACACCTGCAAAATCGTCAATCTGCTGGAGAAAGTGGTTCAGTTGATCGAGCGGTT	534
Db	484	ATTAATTTGCCACAGGTCAAAATTTGTTCTTTTCAGGTCACAAAAGCTTTAAATTGATGAGCTAGTA	543
Qy	535	GAACTTTTGCAGAAGACGAGTGCACAAAGCCTTGATTCCTCTTAAGTGTCAGGTCCCTTT	594
Db	544	GAATAAGGTAAATCATATTAGTGCAAAAGCGTCAATGCCCTTTAGCAGTATCTGGCCCATTC	603
Qy	595	CACACGCTCTCCTTTAGCCTGCTAGCCAGAAACTAGCTGAAACTCTAGCTCAGGTAAGT	654
Db	604	CATTCATCGCTAATGAAAGTGATTGAAGAGATTTTCAAGTTATATTATCAATTTGAA	663
Qy	655	TTTTTCAGATTTTACTGTGTCCTAGTCGGCAATACAGAAGCT---GCTGTGATGCAAAA	711
Db	664	TGCGATTTGATGCTAAGTTTCCCTGTAGTTCAAAATGTAAATGGCAAGGTGAAACTGACAA	723
Qy	712	GAGGACATTTGCTCAGCTCTTGACGGGTGAGTCAAGGAAACCGTTTCGTTTCTATGAAGT	771
Db	724	GAAGTAATTAATCTAATAATGTCGCAACAAATATATTCACGAGTACAAATTCATTACTCA	783
Qy	772	ATTGGGGTCATGCAAGAAGCGCATATAGCAACTTTTATCGAGATTTGACCGGGGAAAGTC	831
Db	784	ACAGAATGGCTAATAGACCAAGGTGTTGATCATTTTATTGAAATTTGGTCTCTGGAAAGTT	843
Qy	832	TTGTCAGGTTTTGTTAAAAAAATTCATCAAACTGCTCACTTAGCTCATGTGGGAAGTCAA	891
Db	844	TTGCTCGGCTTAATTAATAAATAAATATAGAGATTTAGTTTAACATCAATTTCAACTTTA	903
Qy	892	GGCAGTTTAGTACGACTTTTAGAAAA	917
Db	904	GAAGATGTCGAAAGGTGGAATGAAAA	929

RESULT 12

```

RES-09-015-243-4405
Sequence 4405, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.

```

Db 595 CATTTCATCGCTAATGAAGAGTGTGATGAAGAGATTTTCAAGTTATATTAATCAATTTGAA 654  
QY 655 TTTTCAGATTTTACTTGTCCTCCCTAGTCGGCAATACAGAGCT---GCTGTGATGCCAAAAA 711  
Db 655 TGGCATGATGTAAGTTTCCCTGTAGTTCAAAATGTAATGCGCAAGGTGAAGTGAACAA 714  
QY 712 GAGGACATTCGTGAGCTCTTGACGCGTCCAGGTCAAGGAACCCGTTTCGTTTCTATGAAAGT 771  
Db 715 GAAGTAATTAATCTAATATGTCACAAATATTAATCACCAGTACAAATTCATTAATCA 774  
QY 772 ATTGGGTCATGCAAGAGAGGAGGATAGCAATTTATCGAGATGGAACGGGGAAGTC 831  
Db 775 ACAGAAATGGCTAATAGACCAAGGTGTTGATCATTTTATGAAATGCTCTGGAAGTT 834  
QY 832 TTGTCAGGTTTTTGTAAAAAATGATCAAACTGCTCACTTAC 875  
Db 835 TTGTCGCTTAATTAATAAATAATAGAGATGTTAAGTTAAC 878

## RESULT 13

US-09-815-242-7796  
; Sequence 7796, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/205,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7796  
; LENGTH: 939  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(939)  
US-09-815-242-7796

Query Match 13.18; Score 120.8; DB 10; Length 939;  
Best Local Similarity 49.08; Pred. No. 8.3e-28;  
Matches 410; Conservative 0; Mismatches 417; Indels 9; Gaps 3;  
QY 13 GCCTTTTATTTGCTGCTCAAGTGCAGTATCTAGGATGGACGGATTTCTATGAT 72  
Db 16 GCATTCGTTTCCCTGCCAGGTTGCAATCCCTCGCATGCTGCCAGTGGCGCC 75  
QY 73 CAGTATCCGATGTCGAAGAAACGATTGATCGAGCGAGTCAGGTGCTCGTTATGATTTA 132

Db 76 CAGCAGCGCTGGTGGCGGATACCTTCGCGAGGCGCTCGAGGCGCTCGTTACGACCTT 135  
QY 133 CGTTATCTCATCGATACGG---AAGAAGACAACTCAATCAGACCCGCTATACGCAACA 189  
Db 136 TGGGCGCTGGTCCAGATGGTCTGAAGAGCGCTGAACACGACGACAAAGACCCAGCG 195  
QY 190 GCATTCTAGCGACTTCGGTTGCTATCTACCGTTTATTGCAAGAAAGGGCTAT---CAG 246  
Db 196 GCATCTCTACGGTTTCGATCGGCTCTGGCGCTCTGGCTGCGGAGGGCGGTGGCGC 255  
QY 247 CTTGATATGTTTGTGCTGTTTGTCTCTTGGAGAACTCTGCTTGGTGGCAAGCGCGC 306  
Db 256 CCGCGCTTCGTCGCGGCGCACAGCTGGCGCAATATTCGCGCTGTCGCGCCGAAAGC 315  
QY 307 TTGGATTTTGAAGATGCGGTTGCTTGGTAGCTAAGCGTGGAGCCCTATATGAAAGAG 366  
Db 316 CTGGGCTTCGCGCATGCGGTCAAGCTGTCAGCGTAGGGCCAACTGATGACGAGCG 375  
QY 367 GCTCTGTGACTCTGGCAAGATGGTAGCAGTTCTCAATACGCCAGTAGAGTCAATTGA 426  
Db 376 GTTCGCGGGGCGAGGGCGCATGCGCGCATCTTGGCTGGAAGACGCGATGATTG 435  
QY 427 GAAGCTGTCAAAAAGCTTCTGAACCTTGGAGTGGTACTCCAGCCCACTATTAACACCT 486  
Db 436 GCGGCTGTGCGGAGCGGCGCCAGGGCAGGTGGTCAAGCGGTCAATTTCAACGCGCG 495  
QY 487 GCACAAATCGTCATTGCTGGAGAGTGGTTGCAGTTGATCGAGCGTTGAACCTTTTGA 546  
Db 496 GGGCAGGTAGTATCGCGGTGCGCGGCTGCGCTTGGAGCGTGCATCGAGGCATGCAAG 555  
QY 547 GAAGCAGGTGCCAAAGCTTGAATCTCTTAAAGTGTGAGTCCCTTTTACACCCCTCTC 606  
Db 556 GCACGCGCGCAAGCGCGCTTGGCGTTCAGCTGAGCTGCGCTGCGCTGCGCTGCGAACTG 615  
QY 607 CTTGAGCTGTGAGCCAGAACTAGCTGAACTCTAGCTCAGGTAAGTTTTTTCAGATT 666  
Db 616 ATGCTCGGCGCGCGAGAGTTCGCGCTCGTTCGAAAGCTCAGTGGCAGGCGCG 675  
QY 667 ACTTGTCCCTAGTCGGCAATACAGAAGCTGCTGTGATGCAAAAAAGAGACATTCCTCAG 726  
Db 676 AAGATTTGCTGGTGCAGAACGTCAGCGCGCGTGGCGCTGATCTCGATCGGTGGC 735  
QY 727 CTCT---TGACGCGTCAAGTCAAGGAACCCGTTCTGTTCTATGAAGTATGGGCTCATG 783  
Db 736 CCGGACCTGTGGCAGCTGTACAGCCCGTTCGCTGGTGGAGAGCATCCAGCTGCTG 795  
QY 784 CAAGAAGCAGGCATTAAGCAACTTTATCGAGATTGACCGGGGAAAGTCTTGTACAG 839  
Db 796 GCGGAAAGGGGCTCACCGAGCTGGTCCAGTGGCGGCGCGGCAAGGTCTCTGGCAG 851

## RESULT 14

US-09-781-986A-25  
; Sequence 25, Application US/09781986A  
; Publication No. US20030054436A1  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 5255  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/781.986A

```

; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PE248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6251 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-781-986A-25

Query Match      12.9%; Score 118.4; DB 7; Length 6251;
Best Local Similarity 48.8%; Pred. No. 1.6e-26;
Matches 450; Conservative 0; Mismatches 461; Indels 15; Gaps 4;

QY 1 ATGACTAAACAGCCCTTTTATTGCTGCTCAAGTGCCTCCAGTCTCTAGGGATGGGACGG 60
Db 3386 ATGAGTAAACAGCAATTTATTTTCGGGACAAAGTGCCTCCAAAAGTTGGTATGGGCAA 3445

QY 61 GATTCTCATGATCAGTATCGGATGTCGAAGAAACGATTGATCGAGCGAGTCAGTGTCTC 120
Db 3446 GATTCTTTTAAACAAATGATCAAGCAATCTGAAATTTTAACTTCAGCAGCGAACACATTA 3505

QY 121 GGTATGATTACGTTATCTCATCGATAGCGGAAGAGAC---AAACTCAATCAGACCCGC 177
Db 3506 GACTTTGATATTTAGACACAATGTTTACTGATGAAGAGTAAATGGTGAACTGAA 3565

QY 178 TATACCGAACCCCAATTCAGGACTTCGGTGTCTATCTACCGTTTATTCGAAGAAAG 237
Db 3566 AACACACAAACCGACTTATTAGCGCATAGTTCGGCATTA-----TTAGCAGCGTAAAA 3619

QY 238 GGTATCAGCCTCATATGTTGCTGTTGTTGCTCTTGAGAGATCTCTGCCCTTGTGGCA 297
Db 3620 AATTGAATCCTGATTTTACTATGCGGCATAGTTAGGTGAATATCAAGTTTAGTGCA 3679

QY 298 AGCGGCGCTTGATTTTGAAGATCGGTTGCTTGGTAGCTTAAGCTTAAGCTTATATG 357
Db 3680 GCTGACGTATTATCATTTGAAGATGAGTAAATTTGTTAGAAAACGTTGTCATTAATG 3739

QY 358 GAAGAGCGGCTCCTGCTGACTCTGCAAGATGGTAGAGTTCTCAATACGCCAGTAGAG 417
Db 3740 GGCACAGCATTTCTTACTGGTGTAGGAAGCATGGCTGCGAGTATGGGATTAGATTTGAT 3799

QY 418 GTCAATTGAAGAGCCCTGTCA---AAAAGCTTCTGAACCTTGGAGTGGTTACTCCAGCCAAC 474
Db 3800 AAAGTCGATGAAATTTGTAAGTCATTTATCATCTGATGACAAATTAATGAACCAAGCAAC 3859

QY 475 TATAACACACCTGCACAAATCTCATCTGAGAGAGTGGTTGCAAGTATGATGAGCGGTT 534
Db 3860 ATTAATTGCCAGGTCACAAATTTGTTTCAGGTCAAAAGCTTTTAATTGATGAGTAGTA 3919

QY 535 GAACCTTTTCAAGAGACAGGTGCGCAAAACGTTGATTTCTCTTAAAGGTGTCAGGTCCCTTT 594
Db 3920 GAAAAGGTAAATCATTAGTGTGCAAAACGTTGTCGCTTTAGCAGTATCTGGACCAATC 3979

QY 595 CACACCGCTCTCCTTGAGCCTCTAGCCAGAAACCTAGCTGAAACTCTTAGCTCAGGTAAGT 654
Db 3980 CATTCATCGCTAATGAAAGTGAATGAAGAAAGATTTTCAAGTTACATTAATCAATTTGAA 4039

QY 655 TTTTCAGATTTTACTTGTCCCTAGTCGCAATACAGAGCT--GCTGTGATGCAAAA 711
Db 4040 TGGCGTATGCTAAGTTTCTCTAGTTTCAAAATGTAATGCGCAAGGTGAACACTGACAAA 4099

QY 712 GAGGACATTGCTCAGCTCTTGACGCTCAGGTCAAGGAACCCGTTTCGTTTCTATGAAAGT 771
```

```

Db 4100 GAAGTAATTAATCTAATATGTTCAAGCAATTAATTAATCACCAGTACATTAATTAATCA 4159
QY 772 ATTGGGCTCATGCAAGAGCAGGCATTAAGCAACTTTATCGAGATTGACCGGGGAAAGTC 831
Db 4160 ACAGAAATGGCTAATAGACCAGGTTGATCATTTTATTGAATTTGCTCTGGAAAAGTT 4219
QY 832 TTGTCAGGTTTTTGTAAAAAAATGATCAAAACTGCTCACTTAGCTCATGTTGGAAGATCAA 891
Db 4220 TTATCTGGCTTAATTAATAAAAAATAATAGAGATGTTAAGTTAAACATCAATTCAAACTTTA 4279
QY 892 GCGAGTTTAGTAGCACATTTTAGAAAA 917
Db 4280 GAAGATGTGAAGGATGGAATGAAAA 4305

RESULT 15
US-10-331-061-77
; Sequence 77, Application US/10331061
; Publication No. US20030101486A1
; GENERAL INFORMATION:
; APPLICANT: Lassner, Michael
; APPLICANT: Metz, James G
; APPLICANT: Facciotti, Daniel
; TITLE OF INVENTION: SCHIZOCHYTRIUM PKX GENES
; FILE REFERENCE: CGNE.131.02US
; CURRENT APPLICATION NUMBER: US/10/331,061
; CURRENT FILING DATE: 2002-12-27
; PRIOR APPLICATION NUMBER: US/09/231,899
; PRIOR FILING DATE: 1999-01-14
; PRIOR APPLICATION NUMBER: 60/048,650
; PRIOR FILING DATE: 1997-06-04
; PRIOR APPLICATION NUMBER: 09/090,793
; PRIOR FILING DATE: 1998-06-04
; NUMBER OF SEQ ID NOS: 86
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 77
; LENGTH: 7959
; TYPE: DNA
; ORGANISM: Vibrio marinus
; US-10-331-061-77

Query Match      7.5%; Score 69.2; DB 9; Length 7959;
Best Local Similarity 46.8%; Pred. No. 8.7e-11;
Matches 332; Conservative 0; Mismatches 363; Indels 15; Gaps 3;

QY 152 AAGAGACAACTCAATCAGACCGCTATACGCAACCGCCATCTAGCGACTTCGGTTG 211
Db 1991 AAGAGAGCAATTACGTTTAAACGCAACATGCGCAACCGAGTGGTAGTTGAGTGTG 2050

QY 212 CTATCTACCGTTTATTGCAAGAAAAGGGCTATCAGCCTGATATGTTGCTGTTGTTCTC 271
Db 2051 GTCTGTTCAAAACGTTTAAACAGCAGGTTTAAAGCTGATTTTGTCCGGTCATAGTT 2110

QY 272 TTGGAGNATCTCTGCTTGGTGGCAAGCGGCCCTTGGATTTTGAAGATCGGTTGCCT 331
Db 2111 TCGGTGAGTTTAAACCGCATTTATGGGCTGCGGATGTTTGAACGAAACGATTAATGATG 2170

QY 332 TGGTAGTAGCGCTGGAGCCTATATGGAAGAACGGCTCTCTGCTGACTCTGGCAAGATGG 391
Db 2171 TAGCGGTAGTCGTGGTCAAGCAATGGCTGGCCAGAGCAACAAGATTTTGAT-----G 2224

QY 392 TAGCAGTTCTCAATAGCCAGTAGAGTCAATTAAGAAAGCCTGTCAAAAGCTTCTGAAC 451
Db 2225 CAGGTAAAGATGGCGCTGTTGTTGGTATCCAAAGCAAGTCGCTGTGATCATTTGATACCC 2284

QY 452 TTGGAGTGGTTACTCCAGCAACTATAACACACCTCTCACAAATCGTCATTTGCTGGAGAAG 511
Db 2285 TTGATGATGCTCTATTGTTAACTTCACTCGAATACCAAGTTGTTATTGCTGTTACTA 2344

QY 512 TGTTTCAGTTGATCGAGCGGTTGAACTTTTCGAAGACAGGTGCCAACCGCTTGATTC 571
Db 2345 CGGAGCAGGTTGCTGTAGCGGTTTACAACTTAGGTAATGCTGTTTCAAAA---GTTGTGC 2401
```

QY	572	CTCTTAAGGTGTCAGGTCCCTTTTCACACGCTCTCCTTTGAGCCTGCTAGCCAGAACTAG	631
Db	2402	CACTGCCGTATCTGCTGGTTCCATACACCTTTAGTTTCGTCACGCGCAAAACCATTTG	2461
QY	632	CTGAACCTCTAGCTCAGGTAGTCTTTTTCAGATTTTACTTTGTCCTCCTAGTCGGCAATACAG	691
Db	2462	CTAAAGCCGTTGATAGCGCTAAATTTAAAGCGCCCAAGCATTCACAGTGTTCCTAATGGCA	2521
QY	692	AAGCTGCTGTG-----ATGCAAAAAGAGGACATTTGCTCAGCTCTTGACGGGTGAGGTCA	745
Db	2522	CAGGCTTGGTGCATTCAGCAACCGAATGACATTTAGAAAAACCTGAAAAACCATGC	2581
QY	746	AGGAACCCGTTGCTTTCTATGAAAGTATTGGGGTCATGCAAGAAGCAGGCATAAGCAACT	805
Db	2582	TGGAATCTGTTCAATTTCAATCAAGAAATTGACAACTCTATGCTGATGTTGGCGCGGTAT	2641
QY	806	TTATCGAGATTGGACCGGGGAAGTCTTGTGAGGTTTGTGTTAAAAAATT	855
Db	2642	TTATCGAATTTGGTCCAAAGAATGATTAATAAATTGGTTGAAAAACATT	2691

Search completed: June 11, 2003, 20:01:40  
Job time : 171 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 11, 2003, 17:55:46 ; Search time 1424 Seconds

(without alignments)

10474.750 Million cell updates/sec

Title: US-09-308-397-1

Perfect score: 921

Sequence: 1 atgactaaacagcctttt.....tagcacttttagaaaaatag 921

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 segs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estlin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_hcc:\*

9: gb\_estl:\*

10: gb\_est2:\*

11: gb\_hcc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: gb\_gss:\*

18: em\_gss\_hum:\*

19: em\_gss\_inv:\*

20: em\_gss\_pln:\*

21: em\_gss\_prt:\*

22: em\_gss\_fun:\*

23: em\_gss\_mam:\*

24: em\_gss\_mus:\*

25: em\_gss\_other:\*

26: em\_gss\_pro:\*

27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Length	DB ID	Description
C 1	95.6	10.4	1237 17	AF029428
C 2	87.6	9.5	511 14	BQ18656
C 3	82.2	8.9	544 9	AU069821
C 4	53.2	5.8	427 17	AZ049336
C 5	53	5.8	783 14	BQ96704
6	52.6	5.7	491 12	BF512296

7	44	4.8	744	12	BG597049
8	42.6	4.6	414	13	BJ372892
C 9	42.6	4.6	501	13	BM277916
C 10	42.6	4.6	556	13	BM277853
11	42.4	4.6	373	12	BE919981
12	42.4	4.6	435	12	BF188564
13	42.4	4.6	523	12	BF054427
14	42.4	4.6	650	10	BE427401
C 15	42.4	4.6	808	11	AK020696
16	41.8	4.5	450	13	BI130047
17	41.8	4.5	513	13	BI128381
18	41.8	4.5	524	9	AI162058
19	41.8	4.5	531	9	AI161931
C 20	41.8	4.5	711	14	BQ246282
C 21	41.2	4.5	684	14	C25558
C 22	41.2	4.5	714	9	AU060245
C 23	41	4.5	469	12	BF050033
C 24	41	4.5	546	12	BF169335
C 25	41	4.5	575	13	BI500434
C 26	40.8	4.4	484	10	BE605118
C 27	40.8	4.4	534	13	BJ236347
C 28	40.8	4.4	658	9	AI162555
C 29	40.8	4.4	658	12	BG591410
C 30	40.6	4.4	500	13	BM278805
C 31	40.6	4.4	672	9	AU033776
32	40.2	4.4	405	13	BI132128
33	40.2	4.4	425	13	BI129534
34	40.2	4.4	462	13	BI129308
C 35	40.2	4.4	470	10	BE500556
C 36	40.2	4.4	474	10	BE590504
C 37	40.2	4.4	522	13	BQ232102
C 38	40.2	4.4	536	13	BJ236100
C 39	40.2	4.4	555	10	BE607044
C 40	40.2	4.4	556	13	BJ236365
C 41	40.2	4.4	557	10	BE637457
C 42	40.2	4.4	562	10	BE591088
C 43	40.2	4.4	569	10	BE423529
C 44	40.2	4.4	574	14	BQ245331
C 45	40.2	4.4	606	10	BE422581

## ALIGNMENTS

RESULT 1	AF029428/c
LOCUS	AF029428 Salmonella typhimurium LT2, Lambda DASH II Salmonella
DEFINITION	typhimurium genomic clone 1257-T7, DNA sequence.
ACCESSION	AF029428
VERSION	AF029428.1 GI:2570958
KEYWORDS	GSS.
SOURCE	Salmonella typhimurium.
ORGANISM	Salmonella typhimurium.
REFERENCE	1 (bases 1 to 1237)
AUTHORS	Wong,R.M.-Y., Wong,K.K., Benson,N.R. and McClelland,M.
TITLE	Sample sequencing of a Salmonella typhimurium LT2 lambda library: comparison to the Escherichia coli K12 genome
JOURNAL	FEMS Microbiol. Lett. 173 (2), 411-423 (1999)
MEDLINE	99243757
COMMENT	Contact: McClelland M Molecular Biology Sidney Kimmel Cancer Center 3099 Science Park Road, San Diego, CA 92121, USA Email: mclelland@lifsci.sdsu.edu Class: shotgun.

FEATURES	Location/Qualifiers
source	1. 1237
	/organism="Salmonella typhimurium"
	/strain="LT2"
	/db_xref="taxon:602"

```

/clone="1257-T7"
/clone_lib="Salmonella typhimurium LT2, Lambda DASH II"
/notes=vector: Lambda DASH II; sequenced using Li-Cor
sequencer"
BASE COUNT      293 a      345 c      304 g      281 t      14 others
ORIGIN
Query Match      10.4%; Score 95.6; DB 17; Length 1237;
Best Local Similarity 54.7%; Pred. No. 1.1e-17;
Matches 217; Conservative 7; Mismatches 167; Indels 6; Gaps 2;
QY 1 ATGACTAAACAGCCCTTTTATTGCTGCTCAGGTGCCAGTATCTAGGGATGGGACGG 60
Db 398 ATGACGCAATTGCAATTTGTTTCCCGGTTCAGGTTCTCAGAGCGTTGGGATGTTGGCC 339
QY 61 GATTTCTATGATCAGTATCGATGTCAGAAAGCAATGATCGAGCGAGTCAGTGTCTC 120
Db 338 GAGATGGCGCAATATACCTATCGTAGAAGAAACGTTTGTGGAAGCTTCTCGGCTCTG 279
QY 121 GGTATGATTTAGCTTATCTCATCGATACCGA---AGAAAGACAAACTCAATCAGACCCGC 177
Db 278 GGATATGATCTGTGGCGCTCACCCAGCAAGTCCAGCGGAAGTGTATTAACCTTGG 219
QY 178 TATACGCAACAGCCATCTTAGCGACTTGGTGTCTATCTACCGTTTATGCAAGAAAG 237
Db 218 CAGACGCGCGCGCTTATTAAACCGTTCGNTSCGNTWTGCGCTTTTGGCAGCAGCAG 159
QY 238 GGCATATCA---GCTGATATGTTGCTGTTGCTCTCTTGGAGATACCTCCTTGGTG 294
Db 158 GCGGTAAATGCTCGTAAATGGCAGTTCAGCGCTGGCGGATATTTCCGCGCTGTT 99
QY 295 GCAAGCGCGCTTGAATTTGAAGATGGCGTTCCTTGGTACGTAAGCTGGAAGCTAT 354
Db 98 TGMGCTGGCGTCAATCACTTGTGATGCGGTTGCTGCTGTAATGCGCGTAAATTC 39
QY 355 ATGGAAGACGCGCTCTGCTGACTCTGCGAAGATCG 391
Db 38 ATCAGGAACGACCGGAGAACTGGCGCATGTTGG 2

```

```

RESULT 2
BQ818656/c
LOCUS      511 bp      mRNA      linear      EST 01-AUG-2002
DEFINITION 103072D02.y1 C. reinhardtii CC-1690, Deflagellation (normalized),
Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION  BQ818656
VERSION    BQ818656
KEYWORDS   EST.
SOURCE     Chlamydomonas reinhardtii.
ORGANISM   Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadales; Chlamydomonas.
REFERENCE  1 (bases 1 to 511)
AUTHORS   Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre
,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants. Project: 1030
JOURNAL    Unpublished (2002)
COMMENT    Contact: Charles Hauser
          DCMB Box 91000
          Duke University
          Durham, NC 27708-1000
          Tel: 919 613 8159
          Fax: 919 613 8177
          Email: chauser@duke.edu.
FEATURES   source
            1..511
            /organism="Chlamydomonas reinhardtii"
            /strain="CC-1690 wild type mt+ 21gr"
            /db_xref="taxon:3055"
            /clone_lib="C. reinhardtii CC-1690, Deflagellation
(normalized), Lambda Zap II"

```

```

/notes=vector: pBluescript II SK-; Site.1: EcoRI; Site.2:
XhoI; Deflagellation library, constructed by John Davies
and Jeffrey McDermott, combines cDNAs from CC-1690 cells
which had been re-synthesizing flagella for 15, 30 and 60
min after being deflagellated by pH shock. PolyA mRNA was
purified from each sample, pooled and cDNA synthesized.
The cDNA was directionally cloned into lambda Zap II
(Stratagene) in the EcoRI (5') and XhoI (3') sites.
pBluescript II SK- plasmids were excised from the lambda
ZAP clones by superinfection with ExAssist (Stratagene)
phase. The library was normalized using method 4 described
in Bonaldo et al., (1996) Genome Research 6: 791-806."
BASE COUNT      107 a      148 c      126 g      130 t
ORIGIN
Query Match      9.5%; Score 87.6; DB 14; Length 511;
Best Local Similarity 51.3%; Pred. No. 1.7e-15;
Matches 229; Conservative 0; Mismatches 214; Indels 3; Gaps 1;
QY 421 ATTGAAGAAGCGCTGCTCAAAAAGCTTCTGAACCTTGGAGTGGTTACTCCAGCCAACTATAAC 480
Db 511 ATTGCGAAGCGGTGGAAGAGCTGCAAGAGGTGAGTCTGTTTCTCCGTTAACTTTTAAAC 452
QY 481 ACACCTCACAAATCGTCATTGCTGGAGAAAGTGGTTCAGTTGATCGAGCGTTGAACCTT 540
Db 451 TCTCCGGACAGGTGGTTATTTCGGGTCTATAAAGAACGGTGTGAGCGTGTGGCGCTGCC 392
QY 541 TTGCAAGAAGCAGGTGCGCAACGCTTGATTCCTTCTTAAAGTGTGAGTGTCCCTTTCACAC 600
Db 391 TGTAAAGCGGCGGCGCAAAACGCGCTGCGCTTACCAGTGAAGCTGACCTCTCATCTGT 332
QY 601 GCTCTCTTGAAGCGCTGCTAGCCAGAACTAGCTAGTGAACCTCTAGCTCAGGTAAAGTTTCA 660
Db 331 GCGCTGATGAACACAGCAGCGCACAACTGGCAGTAGAATTAGCGAAATCACCTTTAAC 272
QY 661 GATTTTACTTGTCCCTAGTCGSCAATACAGAG---CTGCTGTGATGCAAAAAGAGGAC 717
Db 271 GCACCAACAGTTCCTGTTGTGAATAACGTTGATGTAATGCGAAACCAATGTTGATGCC 212
QY 718 ATTGCTCAGCTCTTTGACGCGTCAGGTCAAGGAACCCGTTCTGTTTCTATGAAGTATTGGG 777
Db 211 ATCCGTGACCACTGTTGACGTAGTGTATTAACCCGTTTCACTGGACGAAGTCTGTTGAG 152
QY 778 GTATGCAAGAAGCAGCGCATTAAGCACTTTATCAGATTGCGACCGGGGAAAGTCTTGCTCA 837
Db 151 TACATGCGCAGCGCAAGCGGTAGAACATCTCTATGAAGTCGCGCGCGGCAAAAGTCTTACT 92
QY 838 GGTTTTCGTTAAAAAAATTGATCAAAAC 863
Db 91 GGCCTGACGAACGCATTTGCGACAC 66

```

```

RESULT 3
AU069821/c
LOCUS      544 bp      mRNA      linear      EST 02-APR-2002
DEFINITION AU069821 Rice panicle at flowering stage Oryza sativa (japonica
cultivar-group) cDNA clone E3726_7A, mRNA sequence.
ACCESSION  AU069821
VERSION    AU069821.1
KEYWORDS   EST.
SOURCE     AU069821.1 GI:5004646
ORGANISM   Oryza sativa (japonica cultivar-group).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE  1 (bases 1 to 544)
AUTHORS   Sasaki,T. and Yamamoto,K.
TITLE     Rice cDNA from panicle at flowering stage
JOURNAL    Unpublished (1996)
COMMENT    Contact: Takuji Sasaki
          National Institute of Agrobiological Resources
          Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
          305-8602, Japan

```

Tel: 81-298-38-7441  
 Fax: 81-298-38-7468  
 Email: tsasaki@abr.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/  
 PROJECT = RGP

FEATURES  
 source Location/Qualifiers  
 1. 544  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="E3726.7A"  
 /clone\_lib="Rice panicle at flowering stage"  
 /dev\_stage="flowering stage"  
 /note="Organ: panicle; Rice cDNA from panicle at flowering stage"  
 BASE COUNT 118 a 161 c 126 g 133 t 6 others  
 ORIGIN

Query Match 8.9%; Score 82.2; DB 9; Length 544;  
 Best Local Similarity 51.4%; Pred. No. 8.3e-14;  
 Matches 280; Conservative 0; Mismatches 259; Indels 6; Gaps 4;  
 QY 316 GAAGATGGGTGCTTGGTGTAGCTAAGCGTGGAGCTATATGGAAGAGCGGCTCCTGCT 375  
 DB 544 GNTGATGGGTGCTGCTGGTGTAGATGCGGCAAGTTCATGCAAGACCGGTACCGAA 485  
 QY 376 GACTGTGCAAGATGGTGTAGCTTCTCAATACGCCAGTAGAGTTC-ATTGAAGAGCGTG 434  
 DB 484 GCGAGGGCGGTATGCGGCAATCATCGGTCTGGAATCATGCGTCTATTGCGAAGCGTG 425  
 QY 435 TCAGAAAGCTTGTGAACCTTGGAGTGGTGTACTCCAGCCCACTATAACACCTGCACAAAT 494  
 DB 424 TNAAGAGCTGCAGAAAGTTCAGGTGTTTCTCCGGTAAACTTTAACTCTCC-GGACAGGT 366  
 QY 495 COTCATCTGTGAGAAGTGGTTCAGATTGATCGAGCGGTGGAACCTTTTGAAGAGCAGG 554  
 DB 365 GGTATTGCCGTATCAAGAGCGTGTGAGCGTGTGCGCTGCTGTAAGCGCGGG 306  
 QY 555 TGCC-AAAGCTGTATCTCTTAAGTGTAGTCCCTTTACACCGCTCTCTCTTACG 613  
 DB 305 GCGNAAAGCGCGGTGCGGTACAGTGTAGCGGTACCGTCTCACTGTGCGGTGATGAAC 246  
 QY 614 CTGTAGCCAGAACTAGCTGAACTCTAGCTCAGGTAAAGTCTTTCAGATTTTACTTGTG 673  
 DB 245 CAGCAGCGCAGAACTGCGAGTAGAATTAGCGAAATCACCCTTAAACGACCACTTC 186  
 QY 674 CCTAGTGGCAATACAGAAG---CTGCTGTATGCAAAAGAGACATTTGCTCAGTCT 730  
 DB 185 CTGTTGTGAATAACGTTGATGTGAATCGGAAACCAATGGTGTATCGCATCGTGACGCAC 126  
 QY 731 TGACGCGTCAGTCAAGAACCCGTTCTGTTTCTATGAAGTATTGGGTCTATGCAAG 790  
 DB 125 TGTACGTGATGTTATACCGGTTTCACTGACGAGAGTCTGTTGAGTACATGCGCGC 66  
 QY 791 CAGGCATTAAGCAACTTTATCGAGATTGACCGGGGAAAGTCTTCTCAGGTTTGTAAAA 850  
 DB 65 AAGCGGTAGACATCTCTATGAAGTCGCGCCCGGCAAGTCTTACTGGCTGACGAAAC 6  
 QY 851 AAATT 855  
 DB 5 GCATT 1

RESULT 4  
 LOCUS AZ049336/c  
 DEFINITION GSSBr0899 Sheared genomic library Brucella melitensis biovar Abortus genomic clone Bc71, DNA sequence.  
 ACCESSION AZ049336  
 VERSION AZ049336.1 GI:7273251  
 KEYWORDS GSS.  
 SOURCE Brucella melitensis biovar Abortus.  
 ORGANISM Brucella melitensis biovar Abortus  
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

REFERENCE  
 AUTHORS Brucellaceae; Brucella.  
 1 (bases 1 to 427)  
 Sanchez,D.O., Zandomeni,R.O., Cravero,S., Verdun,R.E., Pierrou,E., Faccio,P., Diaz,G., Lanzavecchia,S., Aguerro,F., Frasch,A.C.C., Andersson,S.G.E., Rosetti,O.L., Grau,O. and Ugalde,R.A.  
 Gene discovery through genomic sequencing of Brucella abortus Infect. Immun. 69 (2), 865-868 (2001)  
 21101034  
 Contact: Zandomeni, R.  
 Centro de Investigacion en Ciencias Agropecuarias (CICIA)  
 Instituto Nacional de Tecnologia Agropecuaria (INTA) C.C. 25 (1712)  
 Castelar, Buenos Aires, Argentina  
 Tel: 5411-4621-3316/1683  
 Fax: 5411-4481-1316  
 Email: zandomeni@inta.gov.ar  
 Class: shotgun.

FEATURES  
 source Location/Qualifiers  
 1. 427  
 /organism="Brucella melitensis biovar Abortus"  
 /strain="S-2308"  
 /db\_xref="taxon:235"  
 /clone="Bc71"  
 /clone\_lib="Sheared genomic library"  
 /note="Vector: pBluescript SK(-) (STRATAGENE); Genomic DNA was mechanically sheared, blunt ended, and size-fractionated by agarose gel electrophoresis. Fragments between 1.5-3 Kb were recovered and ligated to the pBluescript SK (-) vector."  
 BASE COUNT 86 a 145 c 120 g 74 t 2 others  
 ORIGIN

Query Match 5.8%; Score 53.2; DB 17; Length 427;  
 Best Local Similarity 48.1%; Pred. No. 6e-05;  
 Matches 151; Conservative 0; Mismatches 163; Indels 0; Gaps 0;  
 QY 256 GTTGTCTGTTTCTCTTGGAGAACTCTGCTGTGGCAAGCGGCGCTTGGATTTT 315  
 DB 328 GTGCGCGGTCATTCGGTGTGGCAATATTCGGTCTGTGCGCTGTGCGACGTTTTCATT 269  
 QY 316 GAAGATGCGGTGCGCTTGGTAGCTAAGCGTGGAGCGCTATATGAAGAAGCGGCTCCTGCT 375  
 DB 268 GCGCACGCGCGTCTCTTGGCATTCGCGCAATGCCATGCAGAGCGGTTCGGTC 209  
 QY 376 GACTCTGGCAAGATGTAGCAGTTTCTCAATACGCCAGTAGAGTCAATTGAAGAAGCTGT 435  
 DB 208 GCGCAAGGCGCTATGCGCGCGATCATCGCTCTGGAACACGCGCATGTGGAGCTGTCTGC 149  
 QY 436 CAAAAAGCTTCTGAACCTTGGAGTGGTTACTCCAGCCCACTATAACACACCTGCACAAATC 495  
 DB 148 ATGGAAGCAAGAGCTTCGCGCCGTCAGATCGGCATGACATGGCGCGCCAGCTT 89  
 QY 496 GTCAATTGCTGGAGAGTGGTTTCAGTTGATCGAGCGGTTGAACCTTTTGAAGAAGCAGGT 555  
 DB 88 GTCATATCGGGCGCCAGGCGCGGTTGAACCTGGCGCGAAGCTTGGCTTCTCGAAAAGGCT 29  
 QY 556 GCGAAGCGTTTGT 569  
 DB 28 GCAAGCGCGCCAT 15

RESULT 5  
 LOCUS BQ796704/c  
 DEFINITION EST 5642 Ripening Grape berries Lambda zap II library Vitis vinifera cDNA clone R1051B09 3', mRNA sequence.  
 ACCESSION BQ796704  
 VERSION BQ796704.1 GI:22011670  
 KEYWORDS EST.  
 SOURCE Vitis vinifera.  
 ORGANISM Vitis vinifera.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Vitaceae; Vitis.

```

REFERENCE 1 (bases 1 to 783)
AUTHORS  Abbal.P., Agasse,A., Ageorges,A., Atanassova,R., Barrieu,F.,
          Couture,C., Dedaldechamp,F., Delrot,S., Glissant,D., Grimplet,J.,
          Hamdi,S., Romieu,C. and Terrier.N.
TITLE    Generation of Expressed Sequence Tag from Grape Berry (skin, pulp
          or seeds) at Various Developmental Stages
JOURNAL  Unpublished (2002)
COMMENT  Contact: Romieu C.
          Unite de Recherche des Produits de la Vigne
          Institut National de la Recherche Agronomique
          2, place Viala, 34 060 Montpellier Cedex 01, France
          Tel: 00-33-(0)4-99-61-28-62
          Fax: 00-33-(0)4-99-61-28-57
          Email: romieu@ensam.inra.fr
          Seg primer: T7.

FEATURES             Location/Qualifiers
     source          1..783
                     /organism="Vitis vinifera"
                     /cultivar="Shiraz"
                     /db_xref="taxon:29760"
                     /clone="RT051B09"
                     /clone_lib="Ripening Grape berries Lambda Zap II Library"
                     /dev_stage="ripening stage"
                     /note="Organ: Fruit; Vector: Lambda Zap II; Site 1: Eco RI
                     ; Site 2: XhoI; Oriented library, construction described
                     in Generation of ESTs from Grape Berry (skin, pulp or
                     seeds) at various developmental stages by Terrier,N.,
                     Ageorges,A., Abbal,P., Romieu,C. in J. Plant Physiol. 158
                     (12): 1575-83 2001"

BASE COUNT  220 a 181 c 145 g 237 t
ORIGIN
Query Match      5.8%; Score 53; DB 14; Length 783;
Best Local Similarity 45.8%; Pred. No. 0.0001;
Matches 222; Conservative 0; Mismatches 260; Indels 3; Gaps 1;

QY  425  AAGAGCGCTCTAAAAGCTTCTGAAGTGGAGTGTCTTACTCCAGCACTATAACAC 484
      |||
Db   783  ATGAGCAATGAAGAAGTGTGATGAAGATAATAAGTCCAAATCGCAATTTCTTATGCC 724
      |||
QY  485  CTGCACAAATCGTCATTCGTGGAGAGTGGTTCAGTTCATCGAGCGGTGAACCTTTTCG 544
      |||
Db   723  CTGGGATTTATGAGTCTCTGGAGGAGTGAAGAGATGGAGCGTTGAAGCTAAGGCTA 664
      |||
QY  545  AAGAGCAGGTGCCAACGCTTGATTCCTTTAAGTGTGAGTCCCTTTTCACCGCTC 604
      |||
Db   663  AGTCATTCGAAGGCCGAATGACGGTGGCTTTCAGTGGCTGGTCTTCCACACTCATT 604
      |||
QY  605  TCCTTGAGCTGTAGCCGAACACTAGCTGAACCTCTAGCTAGTGAAGTTTCAGATT 664
      |||
Db   603  TCATGGAACTGTGCTCTCAAGATTGGAAGCCACATGGCAGCTACGGAATCAGAACTC 544
      |||
QY  665  TTACTTGTCCCTAGTGGCAATACAGAAGCTGCTGTGATGCAAAAAGAGGAC---ATTG 721
      |||
Db   543  CAAGAATCCAGTTATATCAAGCTTGATGCACAGCCACATGCAGATCCGACACGATTA 484
      |||
QY  722  CTGAGCTCTTGACGCTCAGGTCAAGAACCCGTTCTGTTCTATGAAGATATGGGGTCA 781
      |||
Db   483  AGAAGATATTTGGCGCGCTCAAGTGAAGTCTCCCGAGTTCTATGGGAAACACAGTGAAGCTC 424
      |||
QY  782  TGAAGAAGCAGGATGAAGCACTTTATCGAGATTGGACCGGGGAAAGTCTTGTCAGGTT 841
      |||
Db   423  TCCTAACCAAGGACTAAAAAGAGTTACGAATTGGGCTGGAAGGTTATCGTGGCA 364
      |||
QY  842  TTGTTAAAAATGATCAAACTGCTCATTAGCTCATGTGGAAGATCAAGCGAGTTTAG 901
      |||
Db   363  TTTTCAAGAGATGGCAAAACTGCCGACTTAGAAGATGTTAGTGTGATTGAGCTGGG 304
      |||
QY  902  TAGCA 906
      |||
Db   303  AATCA 299
      |||

```

RESULT 6  
 BPS12296  
 LOCUS  
 DEFINITION  
 ACCESSION  
 BPS12296  
 VERSION  
 BPS12296.1  
 KEYWORDS  
 EST.  
 SOURCE  
 human.  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 491)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs@mail.nih.gov  
 Oligo-dT track not found, Not 1 site shown in beginning of sequence  
 is likely internal to the message. cDNA Library Preparation: M.B.  
 Soares Lab Clone distribution: NCI-CGAP clone distribution  
 Information can be found through the I.M.A.G.E.B. Consortium/LLNL at:  
 www-bio.llnl.gov/bbrp/image/image.html  
 Seg primer: M13 Forward  
 POLYA-No.

Location/Qualifiers  
 1..491  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3069391"  
 /clone\_lib="NCI CGAP Sub7"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Vector: p773D-pac (Pharmacia) with a modified  
 polylinker; Site 1: Not 1; Site 2: Eco RI; NCI CGAP Sub7  
 is a subtracted library derived from NCI CGAP Sub6. The  
 NCI CGAP Sub7 library had 12 million recombinants. A  
 single-stranded DNA preparation of NCI CGAP Sub6 was used  
 as a tracer in a subtractive hybridization with a driver  
 comprising the IMAGE pool (NCI CGAP Kid3 pool 1 LLAM  
 3334-3337, 3682-3683, 3798-3803 (IMAGE Clones  
 132378-132391, 145608-145675, 150052-150285);  
 NCI CGAP Kids pool 1 LLAM 3338-3342, 3722-3725, 3776-3778  
 (IMAGE Clones 1323912-1325831, 1471368-1472903,  
 1492104-1493255); NCI CGAP LUS pool 1 LLAM 3575-3582,  
 3851-3854 (IMAGE Clones 1414920-1417991, 1520904-1522439  
 ); NCI CGAP GC4 pool 1 LLAM 3164-3167, 3716-3720,  
 3733-3735 (IMAGE Clones 1257096-1258631, 1469064-1470983  
 , 1475592-1476743); NCI CGAP P22 pool 1 LLAM 2457-2459,  
 2758-2759, 3062-3068 (IMAGE Clones 985608-986759,  
 1101192-1101959, 1217928-1220615); NCI CGAP CG10 pool 1  
 LLAM 2644-2653, 2871-2872 (IMAGE Clones 1057416-1061255  
 , 1144584-1145351). (6% of the driver population), plus a  
 pool of 3,840 arrayed clones from NCI CGAP Sub1 (IMAGE  
 Clones 2708616-2710535) and NCI CGAP Sub2 (IMAGE  
 Clones 2710536-2712455) (4% of the driver population  
 ), plus a pool of 11,136 clones from NCI CGAP Sub3 (IMAGE  
 Clones 2712456-2723591) (10% of the driver population),  
 plus a pool of 5,472 clones from NCI CGAP Sub4 (IMAGE  
 Clones 2723592-2729326) (40% of the driver population),  
 plus a pool of 4032 clones from NCI CGAP Sub6 (IMAGE  
 Clones 2728969-2733190) (40% of the driver population).  
 Subtraction was performed as previously described (Bonaldi  
 , Lennon & Soares (1996): Normalization and Subtraction:  
 Two Approaches to Facilitate Gene Discovery. Genome  
 Research 6, 791-806.  
 TAG SEQ=None found"

BASE COUNT 90 a 152 c 167 g 82 t  
 ORIGIN

Query Match 5.7%; Score 52.6; DB 12; Length 491;  
 Best Local Similarity 50.6%; Pred. No. 0.0001;  
 Matches 127; Conservative 0; Mismatches 124; Indels 0; Gaps 0;



BM277916  
 BM277916.1 GI:17971174  
 EST.  
 pig roundworm.  
 Source  
 Ascaris suum  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea  
 ; Ascarididae; Ascaris.  
 1. (bases 1 to 501)  
 Blaxter, M.L., Parkinson, J., Whitton, C., Daub, J., Guiliano, D., Hall  
 N., Quayle, M. and Barrell, B.  
 Edinburgh University/Sanger Centre Nematode EST Project  
 Unpublished (2000)  
 Contact: Blaxter ML  
 Institute of Cell, Animal and Population Biology  
 University of Edinburgh  
 Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9  
 3JT, UK.  
 Tel: +44 131 650 6760  
 Fax: +44 131 670 5450  
 Email: mark.blaxter@ed.ac.uk  
 The library was prepared by Michelle Lizotte-Waniewski for Alan  
 Scott, Johns Hopkins University Medical School, Baltimore.  
 Sequencing was performed by the Pathogen Sequencing Unit, Sanger  
 Centre, Cambridge, UK (Neil Hall, Mike Quail & Bart Barrell).  
 PCR Primers  
 FORWARD: T3  
 BACKWARD: T7PL  
 Plate: 50 row: D column: 12  
 Seq primer: SKPL  
 High quality sequence stop: 446.

```

FEATURES
  source
    1..501
    /organism="Ascaris suum"
    /db_xref="taxon:6353"
    /clone_lib="As tgz_50012"
    /clone_lib="Ascaris suum adult male testis germinal zone
    from Alan Scott"
  "high quality sequence" 446.
  Location/Qualifiers

```

```

187 a 143 C 58 g 113 t
BASE COUNT
/sex="Male"
/dev stage="Adult"
/notes=vector: Lambda Zap II; Site 1: EcoRI; Site 2: XhoI
Library was made from dissected testis germinal zone from
adult male Ascaris suum collected from abattoirs,
Constructed by Michelle Lizotte-Waniewski for Alan Scott,
Johns Hopkins University Medical School, Baltimore, MD."

```

Query Match	4.6%	Score 42.6	DB 13	Length 501
Best Local Similarity	48.2%	Pred. No. 0.12		
Matches 120; Conservative	0	Mismatches 129	Indels 0	Gaps 0

[illegible]

RESULT 10

BM277853/c  
 LOCUS  
 DEFINITION As\_tgz\_49E05 SKPL Ascaris suum adult male testis germinal zone from Alan Scott Ascaris suum cDNA clone As\_tgz\_49E05 5', mRNA sequence.  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Ascaris suum  
 pig roundworm.  
 EST.  
 BM277853.1 GI:11797111  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea  
 1 Ascarididae; Ascaris.  
 1 (bases 1 to 556)  
 AUTHORS Blaxter, M.L., Parkinson, J., Whitton, C., Daub, J., Guiliano, D., Hall, N., Quayle, M. and Barrell, B.  
 TITLE Edinburgh University/Sanger Centre Nematode EST Project  
 JOURNAL  
 Unpublished (2000)  
 COMMENT  
 Institute of Cell, Animal and Population Biology  
 University of Edinburgh  
 Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9  
 3JY, UK.  
 Tel: +44 131 650 6760  
 Fax: +44 131 670 5450  
 Email: mark.blaxter@ed.ac.uk  
 The library was prepared by Michelle Lizotte-Maniewski for Alan Scott, Johns Hopkins University Medical School, Baltimore.  
 Sequencing was performed by the Pathogen Sequencing Unit, Sanger Centre, Cambridge, UK (Neil Hall, Mike Quail & Bart Barrell). The sequence contained a PolyA tail (trimmed)  
 PCR Primers  
 FORWARD: T3  
 BACKWARD: T7PL  
 Plate: 49 row: E column: 05  
 Seq primer: SKPL  
 High quality sequence stop: 542.  
 Location/Qualifiers  
 1..556

```

i. 1. 2506
/organism="Ascaris suum"
/db_xref="taxon:6253"
/clone="As tgz 49E05"
/clone.lib="Ascaris suum adult male testis germinal zone
from Alan Scott"
/sex="Male"
/dev_stage="Adult"
/notes="Vector: Lambda Zap II; Site 1: EcoRI; Site 2: XhoI;
Library was made from dissected testis germinal zone from
adult male Ascaris suum collected from abattoirs.
Constructed by Michelle Lizotte-Waniewski for Alan Scott,
Johns Hopkins University Medical School, Baltimore, MD."
202 a 154 c 76 g 124 t
BASE COUNT

```

BASE COUNT	202 a	154 c	76 g	124 t	ORIGIN
Query Match	4.6%;	Score 42.6;	DB 13;	Length 556;	
Best Local Similarity	48.2%;	Pred. No. 0.13;			
Matches 120;	Conservative 0;	Mismatches 129;	Indels 0;	Gaps 0;	
Qy	308	TGATTTTGAAGANGCGGTTGCGTTGGTAGCTAAGCGTGGAGCGCTATATGGAGAAGACGGG	367		
Db	371	TTCAAGTTGCTGCTGCTGTTGAAGTTGCTGTGCTAAGGTTGTTGAGGTTATTGTTGAGG	312		
Qy	368	CTCCTGCTGACTGCGCAAGATGGGTAGCAGTTTCTCAATACGCCAGTAGAGGTCATTGAAG	427		
Db	311	TTGCTGAGATTGTTGTGAGATAGTTGAGGTTGTTATTGAGGTTGTTGAGGTTGTTCAAG	252		
Qy	428	AAGCCTGTCAAAAGCTTCTGAACCTGGAGTGTTACTCCAGCCCACTATAACACACCTG	487		
Db	251	TTGTAGCTCAGGTTGCTCAGAGATTGCTGAGGTTGTTATTGAAGTTATTTCAGAGAGTTG	192		
Qy	488	CACAAATGCTCATTCGTCGAGAAAGTGTTGCAGTTGATCGACGGGTTGAACTTTTGCAAG	547		
Db	191	CTGAGGTTGTTATTGTAAGTTGCTGTGTTGAAAGTTGCTGTTGAAGTTGCTGTGTTGAAG	132		
Qy	548	AAGCAGGTTG	556		





